EMPIRICAL LIKELIHOOD AND EXTREMES

A Thesis
Presented to
The Academic Faculty

by

Yun Gong

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology
May 2012

EMPIRICAL LIKELIHOOD AND EXTREMES

Approved by:

Professor Liang Peng,
Committee Chair
School of Mathematics
Georgia Institute of Technology

Professor Liang Peng, Advisor
School of Mathematics
Georgia Institute of Technology

Professor Christian Houdré
School of Mathematics
Georgia Institute of Technology

Professor Vladimir Koltchinskii
School of Mathematics
Georgia Institute of Technology

Professor Xiaoming Huo
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Jeff Qin
Department of Mathematics and Statistics
Georgia State University

Date Approved: 16 November 2011

To my parents.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor Liang Peng, without whose mentorship this thesis never would have been written. It has been an honor to be his first Ph.D student. During the past five years, I have greatly enjoyed working with him, and I really appreciate his time and ideas to make my Ph.D study so productive. I am also thankful for the excellent example he provides me as a successful statistician and professor.

I would like to acknowledge my dissertation committee members: Christian Houdré, Xiaoming Huo, Vladimir Koltchinskii and Jeff Qin for their time and insightful comments and questions. I also thank Christian Houdré, Xiaoming Huo and Vladimir Koltchinskii for serving on my oral exam two years ago and discussing with me some possible and interesting directions for my future research work.

For the research on parametric specification test for nested model, I am grateful to Mingqiang Li (formerly a professor in MGT at Georgia Tech) for inspirational discussions with me. He made significant contributions to ideas, the simulation and empirical data analysis. I also would like to thank him for giving me a great introduction to financial derivative markets.

For the smoothed jackknife empirical likelihood method discussed in this thesis, I would like to thank Yongcheng Qi (a professor at University of Minnesota) who pointed out the incompleteness of the proofs for the theorems in the draft and had very helpful discussions with me.

I would like to thank Zhouping Li (a formerly visiting PhD student from China) who I worked jointly on developing empirical likelihood methods for value at risk and intermediate quantiles. I will never forget the enjoyable and fruitful time we had worked together for more than a year.

The School of Mathematics provides very kind and generous helps for me and other international graduate students. Luca Dieci is a great graduate coordinator who helped and advised me on various kinds of problems I have encountered since I lived in Atlanta. I am especially thankful to Cathy Jacobson for giving me an extremely interesting and systematical training on English as well as teaching skills during the first year of the study. Klara Grodzinsky helped me become a good teacher especially during the first two years. I also thank Genola Turner, Sharon McDowell, Christy Dalton and Inetta Worthy for their assistance and conversations for the years.

My life at Georgia Tech was enjoyable partly due to many friends I met here in Atlanta. Huy Huynh is a wonderful friend and roommate. He helped me with so many things since the first day we met, and we had a lot of enjoyable time studying and playing together. I also would like to thank Giang Do, Huijun Feng, Ruoting Gong, Allen Hoffmeyer, Liping Juan, Amey Kaloti, Nan Lu, Jun Lu, Ruodo Wang, Anh Tran, Ralph Yuan as well as many other friends for their helps and wonderful time we spent together.

I am very glad and thankful that Professor Chongchun Zeng recruited me to Georgia Tech. Without his efforts, I would have missed out such a wonderful experience.

Last but not least I would like to thank my parents for their love and encouragement. Without their consistent and patient supports in all my pursuits, I would never have been close to what I am now. Thank you.

TABLE OF CONTENTS

DEDICATION . . . . **iii**

ACKNOWLEDGEMENTS . . . . **iv**

LIST OF TABLES . . . . **viii**

LIST OF FIGURES . . . . **x**

SUMMARY . . . . **xi**

I INTRODUCTION **1**

1.1 Basic concepts and ideas 2

II SMOOTHED JACKKNIFE EMPIRICAL LIKELIHOOD METHOD FOR ROC CURVE **9**

2.1 Introduction 9

2.2 Methodology 12

2.3 Simulation study 14

2.4 Proofs 15

III CONDITIONAL VALUE-AT-RISK IN ARCH/GARCH MODELS 27

3.1 Introduction 27

3.2 Main results 29

3.2.1 Empirical likelihood method for the quantile of ϵ_t 30

3.2.2 Empirical likelihood confidence interval for the conditional VaR 32

3.3 Simulations 33

3.4 Proofs 34

IV CONDITIONAL VALUE-AT-RISK IN HETEROSCEDASTIC REGRESSION MODELS **45**

4.1 Introduction 45

4.2 Main results 47

4.3 Simulation 50

4.4 Proofs 52

V INTERMEDIATE QUANTILES **62**

5.1 Introduction 62

5.2 Methodologies and main results 63

5.3 Simulation study 66

5.4 Proofs 67

VI COVERAGE ACCURACY FOR A MEAN WITHOUT THIRD MOMENT **74**

6.1 Introduction 74

6.2 Main results 76

6.3 Proofs 79

VII A SPECIFICATION TEST FOR NESTED STOCHASTIC MODELS WITH DISCRETE AND DEPENDENT OBSERVATIONS **84**

7.1 Introduction 84

7.2 Parametric specification test for nested models 86

7.2.1 Null hypothesis and assumptions 86

7.2.2 Properties of the empirical process 88

7.2.3 Test statistic and its asymptotic distribution 89

7.3 Simulation study 91

7.4 Empirical tests on CBOE's VIX index 96

7.4.1 Data 96

7.4.2 Asymmetric piece-wise linear stochastic elasticity variance model (ALSEV) 96

7.4.3 Aït-Sahalia's non-linear stochastic elasticity variance model (NLSEV) 97

7.5 Proofs 98

LIST OF TABLES

2.1 Coverage probabilities for the ROC curve $R(0.05)$ are reported for the intervals based on the naive bootstrap method for $R_{m,n}(t)$ (NBM), the proposed jackknife empirical likelihood method (JELM) and the empirical likelihood method (ELM) in Claeskens et al. [11] for levels $\gamma = 0.9, 0.95$ and various sample sizes. 23

2.2 Coverage probabilities for the ROC curve $R(0.1)$ are reported for intervals based on the naive bootstrap method for $R_{m,n}(t)$ (NBM), the proposed jackknife empirical likelihood method (JELM) and the empirical likelihood method (ELM) in Claeskens et al. [11] for levels $\gamma = 0.9, 0.95$ and various sample sizes. 24

2.3 Coverage probabilities for the ROC curve $R(0.25)$ are reported for intervals based on the naive bootstrap method for $R_{m,n}(t)$ (NBM), the proposed jackknife empirical likelihood method (JELM) and the empirical likelihood method (ELM) in Claeskens et al. [11] for levels $\gamma = 0.9, 0.95$ and various sample sizes. 25

2.4 Interval lengths are reported for the ROC curve $R(t)$ based on the naive bootstrap method for $R_{m,n}(t)$ (NBM) and the proposed jackknife empirical likelihood method (JELM) for levels $\gamma = 0.9, 0.95$ and various sample sizes. 26

3.1 Coverage probabilities for the confidence intervals of the conditional VaR based on the proposed empirical likelihood method for levels $\gamma = 0.90, 0.95$. 44

4.1 Coverage probabilities for model (i). 61

4.2 Coverage probabilities for model (ii). 61

5.1 The absolute coverage probability errors for the confidence intervals $I^*_{0.9}(h, n)$ and $I^*_{0.95}(h, n)$ are reported for $q_n = 1 - n^{-a}$ and $h = c\{n(1 - q_n)\}^{-1/3}$. 73

5.2 The absolute coverage probability errors for the confidence intervals $I_{0.9}(h, n)$ and $I_{0.95}(h, n)$ are reported for $q_n = 1 - n^{-a}$, $n = 1000$ and $h = cn^{-1/4}(1 - q_n)^{-\hat{\gamma}(k) - 1/4}$. 73

7.1 Parameter estimates from the parametric specification test and maximum likelihood estimation for nested Vasicek models 110

7.2 Empirical powers and sizes of the parametric specification test and maximum likelihood estimation for nested Vasicek models 110

7.3 MLE and PST for ALSEV on the CBOE VIX Index 112

7.4 MLE and PST for NLSEV on the CBOE VIX Index 112

LIST OF FIGURES

6.1 Plots of $r(x; p, \rho)$ as a function of x and ρ. 83

6.2 Plot of $r^*(x; \rho)$ as a function of x and ρ. 83

7.1 **Comparison of speed of convergence for the two methods of computing $\phi(x, y)$ in the Vasicek model.** The method using Hermite polynomials converges very fast with few summation terms, while the method using bivariate cumulative normal distribution requires hundreds of terms in order to achieve relatively high accuracy. 107

7.2 **The functions $\phi(x, y)$ and $\phi_0(x, y)$ in Theorem 1 for Vasicek model.** The left two subplots are the surface plot (top subplot) and contour plot (bottom subplot) of $\phi(x, y)$, while the right two subplots are those of $\phi_0(x, y)$. The underlying parameters are: $\kappa = 0.1685$, $m = 0.0582$, and $\sigma = 0.0186$. These plots show that autocorrelation terms are very important in $\phi(x, y)$ as the peak of $\phi_0(x, y)$ is only about one tenth of that of $\phi(x, y)$. Also, as the contour plots show, the dependency structures in $\phi(x, y)$ and $\phi_0(x, y)$ are quite different. . . 108

7.3 **Histograms of test statistic from Theorem 1 and from bootstrapping method.** The simulations use the Vasicek model with $\kappa = 0.1685$, $m = 0.0582$, and $\sigma = 0.0186$. The top subplot gives the histogram of the theoretical asymptotic distribution (constant multiple of a chi-squared distribution of degree 1) with 100,000 draws. The bottom subplot is the histogram of the test statistic estimated from 1000 simulations of the Vasicek model. 109

7.4 **Daily CBOE VIX index levels and changes from January 2, 1990 to November 9, 2009.** The top subplot graphs the VIX index levels (normalized to percentage by dividing it by 100). There are a total of 5004 daily observations. A noticeable feature is the high levels during the subprime mortgage crisis in years 2007–2009. The bottom subplot graphs the daily VIX changes. 111

SUMMARY

In 1988, Owen introduced empirical likelihood as a nonparametric method for constructing confidence intervals and regions. Since then, empirical likelihood has been studied extensively in the literature due to its generality and effectiveness. It is well known that empirical likelihood has several attractive advantages comparing to its competitors such as bootstrap: determining the shape of confidence regions automatically using only the data; straightforwardly incorporating side information expressed through constraints; being Bartlett correctable. The main part of this thesis extends the empirical likelihood method to several interesting and important statistical inference situations. This thesis has four components. The first component (Chapter II) proposes a smoothed jackknife empirical likelihood method to construct confidence intervals for the receiver operating characteristic (ROC) curve in order to overcome the computational difficulty when we have nonlinear constrains in the maximization problem. The second component (Chapter III and IV) proposes smoothed empirical likelihood methods to obtain interval estimation for the conditional Value-at-Risk with the volatility model being an ARCH/GARCH model and a nonparametric regression respectively, which have applications in financial risk management. The third component (Chapter V) derives the empirical likelihood for the intermediate quantiles, which plays an important role in the statistics of extremes. Finally, the fourth component (Chapter VI and VII) presents two additional results: in Chapter VI, we present an interesting result by showing that, when the third moment is infinity, we may prefer the Student's t-statistic to the sample mean standardized by the true standard deviation; in Chapter VII, we present a method for testing a subset of parameters for a given parametric model of stationary processes.

CHAPTER I

INTRODUCTION

Empirical likelihood is a nonparametric method for constructing confidence intervals and regions. It was first introduced by Owen (1988, 1990) and has been studied extensively in the literature because of its generality and effectiveness. As an alternative to bootstrap, empirical likelihood has some advantages: for example, it determines the shape of confidence regions automatically using only the data; it straightforwardly incorporates side information expressed through constraints or estimating equations; it is Bartlett correctable, whereas the bootstrap is not, and the coverage accuracy of empirical likelihood can be enhanced by a simple correction (DiCiccio, Hall and Romano 1991).

Likelihood methods are very effective in that they can be used to find efficient estimators and to construct tests with good power properties. Empirical likelihood enjoys the effectiveness of the likelihood approach, as well as the reliability of nonparametric methods such as not having to assume the family of the joint distribution for the data.

Due to these advantages, empirical likelihood has many applications: smooth functions of means (Owen 1990), regression models (Owen 1991), quantile estimation (Chen and Hall 1993), estimating equations (Qin and Lawless 1994), generalized linear models (Kolaczyk 1994), weakly dependent processes (Kitamura 1997), nonparametric density and regression (Chen 1996, Chen and Qin 2003), censored data (Qin and Jing 2001a, Li and Keilegom 2002, Li and Wang 2003), missing data (Wang and Rao 2002), as well as recent rapid and extensive developments in areas such as two sample problems, time series models, longitudinal data, heavy-tailed models, risk models, copulas and tail copulas.

1.1 Basic concepts and ideas

In this section, we will introduce the basic concepts and ideas of empirical likelihood. For more detailed information, see Owen (1988, 1990, 2001).

Consider the independent and identically distributed random variables $X_i \in R^d$, for $i = 1, \ldots, n$, with a common distribution function F_0. Similarly to the parametric likelihood function, its nonparametric counterpart is defined as follows.

Definition 1.1. *Let $X_1, \ldots, X_n \in R^d$ be a random sample with the common distribution function F_0, the nonparametric likelihood function is defined as*

$$L(F) = \prod_{i=1}^{n} F(X_i),$$

where $F(X_i)$ is the probability of getting the observation X_i, $i = 1, \ldots, n$.

Define the empirical distribution function (EDF) of the random sample to be $F_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$. Then the following simple theorem says that F_n is the nonparametric maximum likelihood estimator for the true distribution function F_0.

Theorem 1.1. *Let $X_1, \ldots, X_n \in R^d$ be a random sample with the common distribution function F_0. Let F_n be their EDF and let F be any distribution function. If $F \neq F_n$, then $L(F) < L(F_n)$.*

Suppose we are interested in the statistical inference for a parameter $\theta = T(F)$, where T is a functional of the distributions, $F \in \mathcal{F}$ and $\mathcal{F}$ is taken to be some set of distribution functions. To get an estimator for θ, similarly to the parametric likelihood estimator, we could substitute the distribution function F with its nonparametric maximum likelihood estimator F_n to get $\hat{\theta} = T(F_n)$. Based on the theory of empirical processes and functional delta method, we could obtain the asymptotic limiting distribution for the estimator $\hat{\theta}$, which enables us to construct the approximate confidence interval for the parameter θ. One of the drawbacks of this method for constructing intervals is that one often needs to estimate the unknown parameters appearing in the limiting distribution, which sometimes is difficult and inefficient. Inspired by the parametric likelihood theory, an alternative way to construct confidence interval is based on the likelihood ratio for the parameters and to obtain a nonparametric version of Wilks' theorem. First, we define the nonparametric likelihood ratio function

$$R(F) = \frac{L(F)}{L(F_n)}.$$

And then the profile likelihood ratio function for the parameter θ is defined as

$$\mathcal{R}(\theta) = \sup\{R(F)|T(F) = \theta, F \in \mathcal{F}\}.$$

It can be shown that it is necessary and possible to restrict the distribution function F to distribution functions with support in the sample. Hence, if we let p_i be the weight that F places on observation X_i, where $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$, then the likelihood ratio function becomes

$$R(F) = \prod_{i=1}^{n} np_i.$$

Based on this definition, the rejection region of the hypothesis that $\theta = \theta_0$ has the form $\{\theta|\mathcal{R}(\theta) < r\}$ for some threshold value r_0. And empirical likelihood confidence regions are of the form

$$\{\theta|\mathcal{R}(\theta) \geq r\}.$$

The threshold r is often given the nonparametric analogue of Wilks's theorem, which we will present below. For simplicity, here we outline Owen's empirical likelihood procedure for a mean, which shows the key ideas behind the empirical likelihood method. For more details, see Owen (1990).

Following the above definitions, the profile empirical likelihood ratio function for the mean is defined as

$$\mathcal{R}(\mu) = \sup\{\prod_{i=1}^{n} np_i| \sum_{i=1}^{n} p_i X_i = \mu, p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}. \tag{1.1}$$

Note that in order for (1.1) to have a solution at a candidate μ, μ must be in the convex hull of the sample.

By introducing the Lagrange multipliers $\lambda_0 \in R$ and $\lambda_1 \in R^d$, the constrained optimization problem (1.1) becomes the unconstrained one

$$U(\mathbf{p}, \lambda_0, \lambda_1) = \sum_{i=1}^{n} \log(np_i) + \lambda_0(\sum_{i=1}^{n} p_i - 1) + \lambda_1^T \sum_{i=1}^{n} p_i(X_i - \mu),$$

where $\mathbf{p} = (p_1, \ldots, p_n)^T$. Differentiating $U(\mathbf{p}, \lambda_0, \lambda_1)$ with respect to p_i and let the derivative be zero, we can easily get that $\lambda_0 = -n$ and by defining $\lambda = -n\lambda_1$, we have that the optimal p_i's are given by $p_i = \frac{1}{n}(1 + \lambda^T(X_i - \mu))^{-1}$, where $\lambda \in R^d$ satisfies the following equation

$$\frac{1}{n} \sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda^T(X_i - \mu)} = 0. \tag{1.2}$$

Therefore, by substituting the optimal p_i's into (1.1), the log empirical likelihood ratio becomes

$$l_n(\mu) = -2\log(\mathcal{R}(\mu)) = 2\sum_{i=1}^{n} \log(1 + \lambda^T(X_i - \mu)). \tag{1.3}$$

We have the following nonparametric version of Wilks' theorem.

Theorem 1.2. *(Owen 1990) Let $X_1, \ldots, X_n$ be independent random vectors in R^d with common distribution F_0 having mean μ_0 and finite variance covariance matrix V_0 of rank $q > 0$. Then $l_n(\mu_0)$ converges in distribution to a $\chi^2_{(q)}$ random variable as $n \to \infty$.*

This theorem suggests that we take $r_\alpha = \exp(-\chi^{2,1-\alpha}_{(q)}/2)$ in order to get a $100(1-\alpha)\%$ confidence region $\{\mu | \mathcal{R}(\mu) \geq r_\alpha\}$ for the mean, which can be done by computing $\mathcal{R}(\mu)$ on a grid of μ values and applying a contouring algorithm.

As we can see, in order to evaluate $\mathcal{R}(\mu)$ for each candidate μ, we have to solve a nonlinear system (1.2), which can be computational intensive for $d > 1$. An alternative approach given by Owen (1990) is to transform the optimization problem with respect to p_i's to its dual problem with respect to λ.

Note that the left hand side of (1.2) is the negative gradient with respect to λ of

$$f(\lambda) = -\sum_{i=1}^{n} \log(1 + \lambda^T (X_i - \mu)),$$

and the Hessian of f is

$$H(\lambda) = \sum_{i=1}^{n} \frac{(X_i - \mu)(X_i - \mu)^T}{[1 + \lambda^T (X_i - \mu)]^2}.$$

Moreover, since $0 \le p_i \le 1$ for each i, we can restrict the function f to a convex domain

$$D = \{\lambda : 1 + \lambda^T (X_i - \mu) \ge 1/n, \quad 1 \le i \le n\}.$$

Therefore, on the domain D, $H(\lambda)$ is positive semidefinite and f is the a convex function. The dual problem now involves minimizing a convex function, which can be done using various efficient algorithms from convex programming.

Now, we outline the proof of Theorem 1.2.

Proof. The first key step is to show

$$\|\lambda\| = O_p(n^{-1/2}) \tag{1.4}$$

by expanding (1.2) at μ_0, the true value of μ. Then we can show that

$$\max_{1 \le i \le n} |\lambda^T (X_i - \mu_0)| = o_p(1). \tag{1.5}$$

With (1.5), we are able to invert (1.2) to get

$$\lambda = S^{-1}(\bar{X} - \mu_0) + o_p(n^{-1/2}),$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Expanding (1.3) and using the above expression for λ, we can show that

$$l_n(\mu_0) = n(\bar{X} - \mu_0)^T S^{-1} (\bar{X} - \mu_0) + o_p(1),$$

where $S = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)(X_i - \mu_0)^T$ and Wilks' theorem follows as $S \xrightarrow{p} V_0$ and

$$\sqrt{n}(\bar{X} - \mu_0) \xrightarrow{d} N(0, V_0), \quad \text{as} \quad n \to \infty.$$

□

From the last statement of the proof, we can see that empirical likelihood is based on the normal approximation method. However, to use normal method, we have to estimate the variance V_0, whereas the empirical likelihood method does not require any explicit variance estimation. Owen (1990, 1991) also extended empirical likelihood to smoothed functions of means and linear models.

Qin and Lawless (1994) extended the empirical likelihood for a mean to a more general case of estimating equations, which we discuss briefly as some of the results presented in this thesis are based on the estimating equations.

Again, let $X_1, \ldots, X_n \in R^d$ be a random sample with unknown distribution function F, having a p-dimensional parameter θ. Assume that information about θ and F is available through the following $r \geq p$ estimating equations which are functionally independent,

$$E_F g(x, \theta) = 0,$$

where $g(x, \theta) = (g_1(x, \theta), \ldots, g_r(x, \theta))^T$. As a special case, for the mean of the sample, we can simply choose $g(x, \mu) = x - \mu$, where $\mu \in R^d$ and $r = d$. Now, we can construct empirical likelihood for θ as follows,

$$L(\theta) = \sup\{\prod_{i=1}^{n} p_i | \sum_{i=1}^{n} p_i g(X_i, \theta) = 0, p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}. \tag{1.6}$$

By applying Lagrange multipliers, we get the optimal weight $p_i = \frac{1}{n}(1 + \lambda^T g(X_i, \theta))^{-1}$, where λ satisfies the following equation

$$\frac{1}{n} \sum_{i=1}^{n} \frac{g(X_i, \theta)}{1 + \lambda^T g(X_i, \theta)} = 0.$$

The log empirical likelihood function for θ now becomes

$$l_n(\theta) = -\sum_{i=1}^{n} \log\{1 + \lambda^T(\theta) g(X_i, \theta)\} - n \log n.$$

If $r = p$, then (1.6) is maximized at $\hat{\theta}$, where $\hat{\theta}$ is the solution of

$$\frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta) = 0, \tag{1.7}$$

obtained by substituting the optimal weight $p_i = \frac{1}{n}$ into the estimating equation. Hence, the log likelihood ratio is

$$l_n(\theta) - l_n(\hat{\theta}) = -\sum_{i=1}^{n} \log(1 + \lambda^T(\theta) g(X_i, \theta)),$$

which is the essentially the same as the empirical likelihood for a mean.

In some situations, especially in econometrics models like generalized method of moments, we may have $r > p$, which leads to more estimating equations than the number of parameters in the model. In this case, (1.7) may not have solution. Let $\tilde{\theta}$ maximize the log likelihood function $l_n(\theta)$. Qin and Lawless (1994) showed that under mild conditions, the log likelihood ratio $W_n(\theta_0) = -2\{l_n(\theta_0) - l_n(\tilde{\theta})\}$ for testing $H_0 : \theta = \theta_0$ converges in distribution to χ^2_p as $n \to \infty$, when H_0 is true.

In this thesis, based on the above ideas, we further extend the empirical likelihood to several interesting and important statistical inference situations. The thesis is organized as follows. In chapter II, we propose a smoothed jackknife empirical likelihood method to construct confidence intervals for the receiver operating characteristic (ROC) curve. Jackknife empirical likelihood is first introduced by Jing, Yuan and Zhou (2009) in order to overcome the computational difficulty when we have nonlinear constrains in the maximization problem. From the previous examples, we can see that if the constrains on weights p_i's are linear, after applying the Lagrange multiplier, the problem reduces to solving a fixed number (independent of the sample size n) of equations. However, if the constrains are nonlinear, the Lagrange multiplier may produce a number of equations which may depend on n. Hence, as n gets bigger, the computational difficulty becomes more serious. By applying the standard empirical likelihood method for a mean to the jackknife sample, the empirical likelihood ratio statistic can be calculated by simply solving a single equation. Therefore, this procedure is easy to implement. In chapter III and IV, we propose empirical likelihood methods to obtain interval estimation for the conditional Value-at-Risk with the volatility model being an ARCH/GARCH model and a nonparametric regression respectively. In chapter V, we derive the empirical likelihood for the intermediate quantiles. The results from chapter III, IV and V are useful in financial risk management. Finally, in the last part of the thesis, we present two additional results: in Chapter VI, we present an interesting result by showing that, when the third moment is infinity, we may prefer the Student's t-statistic to the sample mean standardized by the true standard deviation; in Chapter VII, we present a method for testing a subset of parameters for a given parametric model of stationary processes.

CHAPTER II

SMOOTHED JACKKNIFE EMPIRICAL LIKELIHOOD METHOD FOR ROC CURVE

In this chapter we propose a smoothed jackknife empirical likelihood method to construct confidence intervals for the receiver operating characteristic (ROC) curve. By applying the standard empirical likelihood method for a mean to the jackknife sample, the empirical likelihood ratio statistic can be calculated by simply solving a single equation. Therefore, this procedure is easy to implement. The Wilks' theorem for the empirical likelihood ratio statistic is proved and a simulation study is conducted to compare the performance of the proposed method with other methods. The content of this chapter is based on Y. Gong, L. Peng and Y. Qi (2010), Jackknife empirical likelihood method for ROC curve, *Journal of Multivariate Analysis, 101, 1520–1531.*

2.1 Introduction

In diagnostic medicine, it is important to assess the accuracy of a diagnostic test in discriminating diseased patients from non-diseased ones. When the response of a test is continuous, its accuracy is measured by the receiver operating characteristic (ROC) curve; see, e.g., Metz (1978) and Zweig and Campbell (1993). ROC curves can also be used to compare the diagnostic performance of two or more laboratory or diagnostic tests (Griner et al. (1981)).

Let F and G be the distribution functions of the diseased and non-diseased populations, respectively. Then the ROC curve can be written as $R(t) = 1 - F(G^{-}(1-t))$ for $0 < t < 1$, where G^{-} denotes the inverse of G and is defined by $G^{-}(u) = \inf\{x : G(x) \geq u\}$ for $u \in (0, 1)$.

Throughout the chapter, we assume that $X_1, \ldots, X_m$ are independent and identically distributed (i.i.d.) test responses of m patients from the diseased population with distribution F and $Y_1, \ldots, Y_n$ are i.i.d. test responses of n patients from the non-diseased population with distribution G. A simple estimator of $R(t)$ is defined as

$$R_{m,n}(t) = 1 - F_m(G_n^-(1-t)), \tag{2.1}$$

where F_m and G_n are the empirical distribution functions of F and G given by

$$F_m(x) = \frac{1}{m}\sum_{j=1}^{m} I(X_j \le x), \quad G_n(y) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \le y).$$

For the study of the estimator $R_{m,n}(t)$ and its smooth version, we refer to Hsieh and Turnbull (1996), Zhou et al. (1997), Lloyd (1998), Lloyd and Yong (1999), Hall, Hyndman and Fan (2004), Peng and Zhou (2004). For some inference problems related to the ROC curve see, e.g., Gu, Ghosal and Roy (2008) and Zhou (2008).

Using the fact that

$$\sqrt{m+n}\{R_{m,n}(t) - R(t)\} \xrightarrow{d} N(0, \sigma^2(t)), \tag{2.2}$$

where

$$\sigma^2(t) = (1 + \frac{1}{r})R(t)(1 - R(t)) + (1 + r)t(1-t)\{\frac{F'(G^-(1-t))}{G'(G^-(1-t))}\}^2, \tag{2.3}$$

and $r := \lim_{m,n\to\infty} m/n \in (0, \infty)$, one can construct a confidence interval for $R(t)$ via estimating the density functions of F and G or bootstrap methods. As an alternative way to construct confidence intervals without estimating the asymptotic variance explicitly, Claeskens et al. (2003) proposed an empirical likelihood method based on the smoothing estimators of the functions F and G via some link variable. Molanes-Lopez, Van Keilegom and Veraverbeke (2009) studied the empirical likelihood method based on empirical estimators. Qin and Zhou (2006) employed the empirical likelihood method to construct confidence intervals for the area under the ROC curve.

Some recent developments of empirical likelihood methods include inferences for: censored median regression model (Zhao and Chen (2008), Zhao and Yang (2008)), two-sample problems (Liu, Zou and Zhang (2008), Ren (2008), Keziou and Leoni-Aubin (2008), Shen and He (2007), Cao and Van Keilegom (2006), Zhou and Liang (2005)), time series models (Guggenberger and Smith (2008), Chen and Gao (2007), Nordman, Sibbertsen and Lahiri (2007), Nordman and Lahiri (2006), Otsu (2006), Chan and Ling (2006)), longitudinal data and single-index models (Zhao and Jian (2007), Xue and Zhu (2006, 2007), You, Chen and Zhou (2006)) and Copula (Chen, Peng and Zhao (2009)). However, all these applications and extensions of empirical likelihood methods work under linear constraints. In case of nonlinear functionals such as variance, ROC curves and copulas, a common way is to transform nonlinear constraints to linear constraints by introducing some link variables as in Claeskens et al. (2003) and Chen, Peng and Zhao (2009). Unfortunately, this method does not always work and the introduced link variables create more linear constraints, which increases the computational burden. Seeking a general method to deal with nonlinear functionals becomes important.

Recently, Jing, Yuan and Zhou (2009) proposed a so-called jackknife empirical likelihood method for a U-statistic. The procedure is as follows. For a U-statistic, construct a jackknife sample (see, e.g., Shao and Tu (1995)) first, and then treat this jackknife pseudo sample as a sample of i.i.d. observations and apply the standard empirical likelihood method for the mean of i.i.d. observations to obtain the empirical likelihood ratio statistic for the U statistic. Hence, the procedure is easy to implement.

In this chapter, we study the possibility of extending the jackknife empirical likelihood method in Jing, Yuan and Zhou (2009) to construct confidence intervals for the ROC curve so as to avoid adding extra constraints due to the link variable in Claeskens et al. (2003). It turns out that we have to work with a smooth version of the empirical estimator of the ROC curve.

2.2 Methodology

Let w be a symmetric density function with support $[-1,1]$ and put $K(x)=\int_{-\infty}^{x}w(y)dy$. Define the smooth version of $R_{m,n}(t)$ as

$$\hat{R}_{m,n}(t)=1-\frac{1}{m}\sum_{j=1}^{m}K(\frac{1-t-G_n(X_j)}{h}),$$

where $h=h(n)>0$ is a bandwidth. In fact, this smooth estimator of R is obtained via replacing F_m in (2.1) by its smoothed version and G_n is still the empirical distribution of G. Thus, this smoothed estimator of the ROC curve R is different from the one in Claeskens et al. (2003). The reason why we have to work with a smooth version is given in Remark 1 below. Define

$$\hat{R}_{m,n,i}(t)=1-\frac{1}{m-1}\sum_{1\le j\le m, j\ne i}K(\frac{1-t-G_n(X_j)}{h}),\qquad 1\le i\le m,$$

$$\hat{R}_{m,n,i}(t)=1-\frac{1}{m}\sum_{j=1}^{m}K(\frac{1-t-G_{n,i-m}(X_j)}{h}),\qquad m<i\le m+n,$$

where

$$G_{n,k}(y)=\frac{1}{n-1}\sum_{1\le i\le n, i\ne k}I(Y_i\le y),\quad k=1,\ldots,n.$$

The jackknife pseudo sample is therefore defined as

$$\hat{V}_i(t)=(m+n)\hat{R}_{m,n}(t)-(m+n-1)\hat{R}_{m,n,i}(t),\quad i=1,\ldots,m+n.$$

Next, we form the empirical likelihood at $R(t)=\theta$ based on the jackknife pseudo sample as

$$L_{m,n}(t,\theta)=\sup\{\prod_{i=1}^{m+n}p_i : p_1>0,\ldots,p_{m+n}>0,\sum_{i=1}^{m+n}p_i=1,\sum_{i=1}^{m+n}p_i\hat{V}_i(t)=\theta\}.$$

By the standard Lagrange multiplier argument, we obtain that the above maximization is achieved at

$$p_i=\frac{1}{(m+n)\{1+\lambda(\hat{V}_i(t)-\theta)\}},\quad i=1,\ldots,m+n,$$

where $\lambda = \lambda(t, \theta)$ satisfies

$$\frac{1}{m+n} \sum_{i=1}^{m+n} \frac{\hat{V}_i(t) - \theta}{1 + \lambda(\hat{V}_i(t) - \theta)} = 0,$$

which gives the log empirical likelihood ratio as

$$l_{m,n}(t, \theta) = -2 \log L_{m,n}(t, \theta) = 2 \sum_{i=1}^{m+n} \log\{1 + \lambda(\hat{V}_i(t) - \theta)\}.$$

In order to show that the above log empirical likelihood ratio converges in distribution to a χ^2 limit, one has to show that the jackknife variance estimator

$$\nu_{m,n}(t) = \frac{1}{m+n} \sum_{i=1}^{m+n} \{\hat{V}_i(t) - \frac{1}{m+n} \sum_{j=1}^{m+n} \hat{V}_j(t)\}^2$$

is a consistent estimator of $(m+n)Var(\hat{R}_{m,n}(t))$.

Theorem 2.1. *(Gong, Peng and Qi, 2010) Assume that w is a symmetric density with support $[-1, 1]$ and the first derivative of w is bounded. Further assume that the second derivative of $R(t)$ is continuous at $t_0 \in (0, 1)$, and $\lim_{n\to\infty} m/n = r \in (0, \infty)$. If $h = h(n) \to 0$, $nh^2/\log n \to \infty$ and $nh^4 \to 0$ as $n \to \infty$, then*

$$\nu_{m,n}(t_0) \xrightarrow{p} \sigma^2(t_0) \quad as \quad n \to \infty.$$

Remark 2.1. *Although we can not show that the above jackknife variance estimator based on $R_{m,n}(t)$ instead of $\hat{R}_{m,n}(t)$ is inconsistent, our simulation study does confirm this conjecture. This explains why we have to work with a smooth version of the empirical estimator of the ROC curve.*

Theorem 2.2. *(Gong, Peng and Qi, 2010) Under the conditions of Theorem 2.1, we have*

$$l_{m,n}(t_0, R(t_0)) \xrightarrow{d} \chi^2(1) \quad as \quad n \to \infty.$$

Based on Theorem 2.2, a confidence interval with level γ for $R(t_0)$ can be constructed as

$$I_\gamma(t_0, m, n) = \{\theta : l_{m,n}(t_0, \theta) \le \chi^2_{1,\gamma}\},$$

where $\chi^2_{1,\gamma}$ is the γ quantile of $\chi^2(1)$.

2.3 Simulation study

In this section, we compare the coverage accuracy of the proposed jackknife empirical likelihood method with the normal approximation method and the empirical likelihood method in Claeskens et al. (2003), where an extra constraint and smooth distribution estimation for both populations are required.

We consider the following three cases:

$$A)F \sim N(0,1), G \sim N(1,0.5); B)F \sim N(0,1), G \sim Exp(1); C)F \sim Exp(1), G \sim Exp(1),$$

where $Exp(1)$ denotes the standard exponential distribution function. We generate 10,000 random samples from the above cases with sample sizes $m = 50, 100, 200$ and $n = 50, 100, 200$. We use the kernel $w(x) = \frac{15}{16}(1-t^2)^2 I(|t| \leq 1)$ for both methods, $h = m^{-1/3}$ for the jackknife empirical likelihood method and $h_1 = m^{-1/3}$ and $h_2 = n^{-1/3}$ for the empirical likelihood method in Claeskens et al. (2003). Note that Chen, Peng and Zhao (2009) pointed out that the above choices of bandwidth for the method in Claeskens et al. (2003) are valid. For the naive bootstrap method based on $R_{m,n}(t)$, we employ 1000 bootstrap samples. We compute the coverage probabilities for $t_0 = 0.05, 0.10, 0.25$ with confidence levels 0.9 and 0.95. From Tables 2.1-2.3, we observe that both the proposed jackknife empirical likelihood method and the empirical likelihood method in Claeskens et al. (2003) perform much better than the naive bootstrap method. When $t = 0.05$ and 0.10, the proposed jackknife empirical likelihood method performs best in most cases. Both empirical lieklihood methods are comparable in case of $t = 0.25$. However, the proposed jackknife empirical likelihood method is less computationally intensive since the empirical likelihood method in Claeskens et al. (2003) has more constraints in the optimization procedure. Indeed, we employ the "emplik" R package for the proposed jackknife empirical likelihood method.

Next we examine the interval lengths of the proposed jackknife empirical likelihood method and the naive bootstrap method based on $R_{m,n}(t)$ since the computation for the other empirical likelihood interval is quite intensive. Note that $l_{m,n}(t, \frac{1}{m+n}\hat{V}_i(t)) = 0$. So we both increase and decrease θ from $\frac{1}{m+n}\hat{V}_i(t)$ with a step 0.001 till $l_{m,n}(t,\theta) > \chi^2_{1,\gamma}$ to obtain the upper and lower endpoint of the jackknife empirical likelihood interval $I_\gamma(t_0, m, n)$. In Table 2.4, we report the interval lengths for the jackknife empirical likelihood method and the naive bootstrap method. We observe that the jackknife empirical likelihood method results in a shorter interval than the naive bootstrap method for cases A, B and C with $\gamma = 0.9$. But, for case C with $\gamma = 0.95$, the jackknife empirical likelihood method produces a longer interval than the naive bootstrap method.

2.4 Proofs

We need the following lemmas to prove Theorems 2.1 and 2.2.

Lemma 2.1. *Assume conditions in Theorem 2.1 hold. Then there exists an interval $(a,b) \subset (0,1)$ such that $t_0 \in (a,b)$ and*

$$\sqrt{m+n}\{\hat{R}_{m,n}(t) - R(t)\} \xrightarrow{D} \sqrt{1+\frac{1}{r}}B_1(1-R(t)) + \sqrt{1+r}R'(t)B_2(t) \tag{2.4}$$

in $D((a,b))$, where $B_1(t)$ and $B_2(t)$ are two independent Brownian bridges.

Proof. Since R'' is continuous at $t_0 \in (0,1)$, there exists a subset (a,b) containing t_0 such that R' and R'' are bounded in (a,b). It is known that

$$\sqrt{m}\{F_m(x) - F(x)\} \xrightarrow{D} W_1(x) \quad \text{and} \quad \sqrt{n}\{G_n(y) - G(y)\} \xrightarrow{D} W_2(y) \tag{2.5}$$

in $D((-\infty,\infty))$, where W_1 and W_2 are two independent Wiener processes with zero means and covariances

$$\begin{cases} EW_1(x_1)W_1(x_2) = F(x_1 \wedge x_2) - F(x_1)F(x_2) \\ EW_2(y_1)W_2(y_2) = G(y_1 \wedge y_2) - G(y_1)G(y_2). \end{cases}$$

Write

$$
\begin{aligned}
& 1-\tfrac{1}{m}\textstyle\sum_{j=1}^{m} K(\tfrac{1-t-G(X_j)}{h})-R(t) \\
= \;& F(G^{-}(1-t))-\textstyle\int_{-\infty}^{\infty} K(\tfrac{1-t-G(x)}{h})\,dF_m(x) \\
= \;& F(G^{-}(1-t))-\textstyle\int_{-\infty}^{\infty} F_m(x)w(\tfrac{1-t-G(x)}{h})h^{-1}\,dG(x) \\
= \;& F(G^{-}(1-t))-\textstyle\int_{-1}^{1} F_m(G^{-}(1-t-xh))w(x)\,dx \\
= \;& F(G^{-}(1-t))-F_m(G^{-}(1-t)) \\
& -\textstyle\int_{-1}^{1}\{F(G^{-}(1-t-xh))-F(G^{-}(1-t))\}w(x)\,dx \\
& -\textstyle\int_{-1}^{1}\{F_m(G^{-}(1-t-xh))-F(G^{-}(1-t-xh)) \\
& -F_m(G^{-}(1-t))+F(G^{-}(1-t))\}w(x)\,dx
\end{aligned}
\tag{2.6}
$$

and

$$
\begin{aligned}
& \textstyle\int_{-1}^{1}\{F(G^{-}(1-t-xh))-F(G^{-}(1-t))\}w(x)\,dx \\
= \;& -\textstyle\int_{-1}^{1} R'(t)xhw(x)\,dx-\tfrac{1}{2}\textstyle\int_{-1}^{1} R''(t^*)(xh)^2w(x)\,dx \\
= \;& -\tfrac{1}{2}h^2\textstyle\int_{-1}^{1} R''(t^*)x^2w(x)\,dx,
\end{aligned}
\tag{2.7}
$$

where t^* is between t and $t+xh$. If follows from conditions in Lemma 2.1 and (2.7) that

$$
\int_{-1}^{1}\{F(G^{-}(1-t-xh))-F(G^{-}(1-t))\}w(x)\,dx=O(h^2) \tag{2.8}
$$

uniformly in $t\in(a,b)$. Using the conditions on h, (2.5) and the continuity of W_1, we have

$$
\begin{aligned}
& \textstyle\int_{-1}^{1}\{F_m(G^{-}(1-t-xh))-F(G^{-}(1-t-xh)) \\
& -F_m(G^{-}(1-t))+F(G^{-}(1-t))\}w(x)\,dx \\
= \;& \textstyle\int_{-1}^{1}\{F_m(G^{-}(1-t-xh))-F(G^{-}(1-t-xh)) \\
& -m^{-1/2}W_1(G^{-}(1-t-xh))\}w(x)\,dx \\
& -\textstyle\int_{-1}^{1}\{F_m(G^{-}(1-t))-F(G^{-}(1-t))-m^{-1/2}W_1(G^{-}(1-t))\}w(x)\,dx \\
& +\textstyle\int_{-1}^{1}\{m^{-1/2}W_1(G^{-}(1-t-xh))-m^{-1/2}W_1(G^{-}(1-t))\}w(x)\,dx \\
= \;& o_p(m^{-1/2}).
\end{aligned}
$$

Hence

$$
\sqrt{m}\{1-\frac{1}{m}\sum_{j=1}^{m}K(\frac{1-t-G(X_j)}{h})-R(t)\}\xrightarrow{D}W_1(G^{-}(1-t)) \tag{2.9}
$$

in $D((a,b))$.

Write

$$
\begin{aligned}
& \tfrac{1}{m}\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_j)}{h}) - \tfrac{1}{m}\sum_{j=1}^m K(\tfrac{1-t-G(X_j)}{h}) \\
= \; & \tfrac{1}{m}\textstyle\sum_{j=1}^m \tfrac{G(X_j)-G_n(X_j)}{h} w(\tfrac{1-t-G(X_j)}{h}) \qquad (2.10)\\
& +\tfrac{1}{2m}\textstyle\sum_{j=1}^m (\tfrac{G(X_j)-G_n(X_j)}{h})^2 w'(\tfrac{1-t-G(X_j)+\xi_{n,j}}{h}),
\end{aligned}
$$

where $\xi_{n,j}$ is between 0 and $G(X_j)-G_n(X_j)$. It follows from Theorem A of Silverman (1978) that

$$
\sup_{t\in(a,b)} |\frac{1}{mh}\sum_{j=1}^m |w'(\frac{1-t-G(X_j)}{h})| - R'(t)\int_{-1}^1 |w'(x)|\,dx| = o_p(1), \qquad (2.11)
$$

where $R'(1-x)$ is the density of $G(X_1)$. By (2.5), (2.10) and (2.11), we have

$$
\begin{aligned}
& \sqrt{n}\{\tfrac{1}{m}\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_j)}{h}) - \tfrac{1}{m}\sum_{j=1}^m K(\tfrac{1-t-G(X_j)}{h})\} \\
= \; & -\textstyle\int_{-\infty}^{\infty} W_2(x)h^{-1}w(\tfrac{1-t-G(x)}{h})\,dF(x) + O_p(n^{-1/2}h^{-1}) \qquad (2.12)\\
= \; & \textstyle\int_{-1}^{1} W_2(G^-(1-t-xh))h^{-1}w(x)\,dF(G^-(1-t-hx)) + O_p(n^{-1/2}h^{-1}) \\
= \; & -R'(t)W_2(G^-(1-t)) + o_p(1)
\end{aligned}
$$

uniformly in $t\in(a,b)$. Hence the lemma follows from (2.9) and (2.12) with $B_1(1-R(t)) = W_1(G^-(1-t))$ and $B_2(t) = W_2(G^-(1-t))$. This completes the proof of the lemma.

□

Lemma 2.2. *Under conditions of Theorem 2.1, we have*

$$
\sqrt{m+n}\{\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(t) - R(t)\} \xrightarrow{d} N(0,\sigma^2(t))
$$

as $n\to\infty$ for $t=t_0$.

Proof. Throughout we assume $t=t_0$. It follows from the definition of $\hat{V}_i(t)$ that

$$
\begin{aligned}
& \tfrac{1}{m+n}\textstyle\sum_{i=1}^{m+n}\hat{V}_i(t) \\
= \; & \tfrac{1}{m+n}\{m+n-\tfrac{m+n}{m}\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_j)}{h}) \qquad (2.13)\\
& +\tfrac{m+n-1}{m}\textstyle\sum_{k=1}^n\sum_{j=1}^m\{K(\tfrac{1-t-G_{n,k}(X_j)}{h}) - K(\tfrac{1-t-G_n(X_j)}{h})\}\}.
\end{aligned}
$$

Write

$$\begin{aligned}
& \textstyle\sum_{k=1}^{n}\sum_{j=1}^{m}\{K(\frac{1-t-G_{n,k}(X_j)}{h})-K(\frac{1-t-G_n(X_j)}{h})\} \\
= \ & \textstyle\sum_{k=1}^{n}\sum_{j=1}^{m}\frac{G_n(X_j)-G_{n,k}(X_j)}{h}w(\frac{1-t-G_n(X_j)}{h}) \\
& +\textstyle\sum_{k=1}^{n}\sum_{j=1}^{m}\frac{1}{2}\{\frac{G_n(X_j)-G_{n,k}(X_j)}{h}\}^2 w'(\frac{1-t-\xi_{n,k,j}}{h}) \\
= \ & \textstyle\sum_{j=1}^{m}\{\sum_{k=1}^{n}\frac{G_n(X_j)-G_{n,k}(X_j)}{h}\}w(\frac{1-t-G_n(X_j)}{h}) \\
& +\textstyle\sum_{k=1}^{n}\sum_{j=1}^{m}\frac{1}{2}\{\frac{G_n(X_j)-G_{n,k}(X_j)}{h}\}^2 w'(\frac{1-t-\xi_{n,k,j}}{h}) \\
= \ & \textstyle\sum_{k=1}^{n}\sum_{j=1}^{m}\frac{1}{2}\{\frac{G_n(X_j)-G_{n,k}(X_j)}{h}\}^2 w'(\frac{1-t-\xi_{n,k,j}}{h}),
\end{aligned} \tag{2.14}$$

where $\xi_{n,k,j}$ is a random variable between $G_{n,k}(X_j)$ and $G_n(X_j)$. Since

$$G_n(X_j)-G_{n,k}(X_j)=\frac{1}{n-1}\{G_n(X_j)-I(Y_k\le X_j)\}=O_p(\frac{1}{n-1})$$

uniformly in $1\le k\le n$ and $1\le j\le m$, we can write

$$\xi_{n,k,j}=G_n(X_j)+O_p(\frac{1}{n-1})=G(X_j)+O_p(n^{-\frac{1}{2}}). \tag{2.15}$$

It follows from (2.14), (2.15) and (2.11) that

$$\sum_{k=1}^{n}\sum_{j=1}^{m}\{K(\frac{1-t-G_{n,k}(X_j)}{h})-K(\frac{1-t-G_n(X_j)}{h})\}=O_p\{\frac{mn}{(n-1)^2h}\}. \tag{2.16}$$

By (2.13), (2.16) and Lemma 2.1, we have

$$\begin{aligned}
& \sqrt{m+n}\{\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(t)-R(t)\} \\
= \ & \sqrt{m+n}\{1-\frac{1}{m}\sum_{j=1}^{m}K(\frac{1-t-G_n(X_j)}{h})+O_p\{\frac{(m+n-1)n}{(m+n)(n-1)^2h}\}-R(t)\} \\
= \ & \sqrt{m+n}\{\hat{R}_{m,n}(t)-R(t)+O_p\{\frac{(m+n-1)n}{(m+n)(n-1)^2h}\}\} \\
\xrightarrow{d} \ & N(0,\sigma^2(t)),
\end{aligned}$$

i.e., Lemma 2.2 holds. □

Lemma 2.3. *Under conditions of Theorem 2.1, we have*

$$\frac{1}{m+n}\sum_{i=1}^{m+n}\{\hat{V}_i(t)-R(t)\}^2\xrightarrow{p}\sigma^2(t)$$

as $n\to\infty$ *for* $t=t_0$.

Proof. Throughout we assume $t = t_0$. For $1 \le i \le m$, we can write that

$$\hat{V}_i(t) = 1 + \frac{n}{(m-1)m}\sum_{j=1}^{m} K(\frac{1-t-G_n(X_j)}{h}) - \frac{m+n-1}{m-1}K(\frac{1-t-G_n(X_i)}{h})$$

and

$$\begin{aligned}
&\hat{V}_i^2(t) \\
= \; & \{1 - \tfrac{m+n-1}{m-1}K(\tfrac{1-t-G_n(X_i)}{h})\}^2 + \{\tfrac{n}{(m-1)m}\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_j)}{h})\}^2 \\
& +2\{\tfrac{n}{(m-1)m}\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_j)}{h})\}\{1 - \tfrac{m+n-1}{m-1}K(\tfrac{1-t-G_n(X_i)}{h})\},
\end{aligned}$$

which imply that

$$\begin{aligned}
&\textstyle\sum_{i=1}^m \hat{V}_i^2(t) \\
= \; & m - \tfrac{2(m+n-1)}{m-1}\textstyle\sum_{i=1}^m K(\tfrac{1-t-G_n(X_i)}{h}) + \tfrac{(m+n-1)^2}{(m-1)^2}\textstyle\sum_{i=1}^m K^2(\tfrac{1-t-G_n(X_i)}{h}) \\
& +\tfrac{mn^2}{(m-1)^2m^2}\{\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_j)}{h})\}^2 \\
& +\tfrac{2n}{(m-1)m}\{\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_j)}{h})\}\{m - \tfrac{m+n-1}{m-1}\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_i)}{h})\}.
\end{aligned} \tag{2.17}$$

Since K^2 is a distribution function, it follows from Lemma 2.1 that

$$\frac{1}{m}\sum_{i=1}^{m} K^2(\frac{1-t-G_n(X_i)}{h}) \xrightarrow{p} F(G^-(1-t)). \tag{2.18}$$

Hence, by (2.17), (2.18) and Lemma 2.1,

$$\begin{aligned}
&\tfrac{1}{m+n}\textstyle\sum_{i=1}^m \hat{V}_i^2(t) \\
\xrightarrow{p} \; & \tfrac{r}{1+r} - 2F(G^-(1-t)) + (1+\tfrac{1}{r})F(G^-(1-t)) \\
& +\tfrac{1}{r(1+r)}F^2(G^-(1-t)) + \tfrac{2}{1+r}F(G^-(1-t)) - \tfrac{2}{r}F^2(G^-(1-t)) \\
= \; & \tfrac{r}{1+r} + \tfrac{1+2r-r^2}{r(1+r)}F(G^-(1-t)) - \tfrac{1+2r}{r(1+r)}F^2(G^-(1-t)) \\
= \; & \tfrac{r+1}{r}R(t) - \tfrac{1+2r}{r(1+r)}R^2(t).
\end{aligned} \tag{2.19}$$

Next, for $m < i \le m+n$, we can write that

$$\begin{aligned}
&\hat{V}_i(t) \\
= \; & 1 - \tfrac{1}{m}\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_j)}{h}) \\
& +\tfrac{m+n-1}{m}\textstyle\sum_{j=1}^m\{K(\tfrac{1-t-G_{n,i-m}(X_j)}{h}) - K(\tfrac{1-t-G_n(X_j)}{h})\}
\end{aligned}$$

and

$$
\begin{aligned}
& \hat V_i^2(t) \\
= \quad & \{1-\tfrac{1}{m}\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_j)}{h})\}^2 \\
& +\{\tfrac{m+n-1}{m}\textstyle\sum_{j=1}^m\{K(\tfrac{1-t-G_{n,i-m}(X_j)}{h})-K(\tfrac{1-t-G_n(X_j)}{h})\}\}^2 \\
& +2\{1-\tfrac{1}{m}\textstyle\sum_{j=1}^m K(\tfrac{1-t-G_n(X_j)}{h})\} \\
& \times\tfrac{m+n-1}{m}\textstyle\sum_{j=1}^m\{K(\tfrac{1-t-G_{n,i-m}(X_j)}{h})-K(\tfrac{1-t-G_n(X_j)}{h})\}.
\end{aligned}
\tag{2.20}
$$

It follows from (2.11) that

$$
\begin{aligned}
A_k \; & := \{\textstyle\sum_{j=1}^m\{K(\tfrac{1-t-G_{n,k}(X_j)}{h})-K(\tfrac{1-t-G_n(X_j)}{h})\}\}^2 \\
& = \{\textstyle\sum_{j=1}^m \tfrac{G_n(X_j)-G_{n,k}(X_j)}{h} w(\tfrac{1-t-G_n(X_j)}{h}) \\
& \quad + \textstyle\sum_{j=1}^m \tfrac{\{G_n(X_j)-G_{n,k}(X_j)\}^2}{2h^2} w'(\tfrac{1-t-\xi_{n,k,j}}{h})\}^2 \\
& = \{\textstyle\sum_{j=1}^m \tfrac{G_n(X_j)-G_{n,k}(X_j)}{h} w(\tfrac{1-t-G_n(X_j)}{h}) + O_p(mn^{-2}h^{-1})\}^2 \\
& = \{\textstyle\sum_{j=1}^m \tfrac{G_n(X_j)-G_{n,k}(X_j)}{h} w(\tfrac{1-t-G_n(X_j)}{h})\}^2 + O_p(n^{-1}h^{-1}),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
& \tfrac{1}{m+n}\textstyle\sum_{k=1}^n A_k \\
= \quad & \tfrac{1}{m+n}\textstyle\sum_{k=1}^n\{\sum_{l=1}^m\sum_{j=1}^m \tfrac{G_n(X_l)-G_{n,k}(X_l)}{h}\tfrac{G_n(X_j)-G_{n,k}(X_j)}{h} \\
& \times w(\tfrac{1-t-G_n(X_l)}{h})w(\tfrac{1-t-G_n(X_j)}{h})\} + O_p(n^{-1}h^{-1}) \\
= \quad & \tfrac{1}{m+n}\tfrac{n}{(n-1)^2h^2}\textstyle\sum_{l=1}^m\sum_{j=1}^m\{G_n(X_l\wedge X_j)-G_n(X_l)G_n(X_j)\} \\
& \times w(\tfrac{1-t-G_n(X_l)}{h})w(\tfrac{1-t-G_n(X_j)}{h}) + O_p(n^{-1}h^{-1}) \\
= \quad & \tfrac{1}{m+n}\tfrac{n}{(n-1)^2h^2}\textstyle\sum_{l=1}^m\sum_{j=1}^m\{G(X_l\wedge X_j)-G(X_l)G(X_j)\} \\
& \times w(\tfrac{1-t-G(X_l)}{h})w(\tfrac{1-t-G(X_j)}{h})\{1+o_p(1)\} + O_p(n^{-1}h^{-1}) \\
\xrightarrow{p} \quad & \tfrac{r^2}{1+r}\{1-t-(1-t)^2\}\{R'(t)\}^2 \\
= \quad & \tfrac{r^2}{1+r}t(1-t)\{R'(t)\}^2.
\end{aligned}
\tag{2.21}
$$

By (2.20), (2.21), (2.16) and Lemma 2.1, we have

$$
\frac{1}{m+n}\sum_{i=m+1}^{m+n}\hat V_i^2(t) \xrightarrow{p} \frac{1}{1+r}R^2(t)+(r+1)t(1-t)\{R'(t)\}^2. \tag{2.22}
$$

Hence, it follows from (2.19), (2.22) and Lemma 2.2 that

$$\begin{aligned}
&\frac{1}{m+n}\sum_{i=1}^{m+n}\{\hat{V}_i(t)-R(t)\}^2 \\
= &\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i^2(t)+R^2(t)-\frac{2}{m+n}R(t)\sum_{i=1}^{m+n}\hat{V}_i(t) \\
\xrightarrow{p} &\sigma^2(t).
\end{aligned}$$

This completes the proof of Lemma 2.3. $\square$

Proof of Theorem 2.1. It follows immediately from Lemmas 2.2 and 2.3. $\square$

Proof of Theorem 2.2. Throughout let $\theta = R(t_0)$. Define $g(\lambda) = \frac{1}{m+n}\sum_{i=1}^{m+n}\frac{\hat{V}_i(t_0)-\theta}{1+\lambda(\hat{V}_i(t_0)-\theta)}$. It is easy to check that

$$\begin{aligned}
0 = |g(\lambda)| &= \frac{1}{m+n}\left|\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta)-\lambda\sum_{i=1}^{m+n}\frac{(\hat{V}_i(t_0)-\theta)^2}{1+\lambda(\hat{V}_i(t_0)-\theta)}\right| \\
&\geq \left|\frac{\lambda}{m+n}\sum_{i=1}^{m+n}\frac{(\hat{V}_i(t_0)-\theta)^2}{1+\lambda(\hat{V}_i(t_0)-\theta)}\right|-\left|\frac{1}{m+n}\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta)\right| \\
&\geq \frac{|\lambda|S_{m+n}}{1+|\lambda|Z_{m+n}}-\left|\frac{1}{m+n}\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta)\right|,
\end{aligned}$$

where $S_{m+n} = \frac{1}{m+n}\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta)^2$ and $Z_{m+n} = \max_{1\leq i\leq m+n}|\hat{V}_i(t_0)-\theta|$. Using similar arguments in proving Lemma 2.2, we can show that Z_{m+n} is bounded in probability. Hence, by Lemma 2.2, Lemma 2.3 and the fact that Z_{m+n} is bounded in probability, we have

$$|\lambda| = O_p\{(m+n)^{-\frac{1}{2}}\}. \tag{2.23}$$

Put $\gamma_i = \lambda(\hat{V}_i(t_0)-\theta)$. Then, we have that

$$\max_{1\leq i\leq m+n}|\gamma_i| = o_p(1). \tag{2.24}$$

Using (2.23), (2.24) and Taylor expansion, we have

$$\begin{aligned}
0 = g(\lambda) &= \frac{1}{m+n}\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta)(1-\gamma_i+\frac{\gamma_i^2}{1+\gamma_i}) \\
&= \frac{1}{m+n}\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta) - S_{m+n}\lambda + \frac{1}{m+n}\sum_{i=1}^{m+n}\frac{(\hat{V}_i(t_0)-\theta)\gamma_i^2}{1+\gamma_i} \\
&= \frac{1}{m+n}\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta) - S_{m+n}\lambda + O_p(\frac{1}{m+n}),
\end{aligned}$$

which implies that

$$\lambda = S_{m+n}^{-1}\frac{1}{m+n}\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta) + \beta_n, \tag{2.25}$$

where $\beta_n = O_p(\frac{1}{m+n})$. Hence, it follows from (2.23), (2.25), Lemmas 2.2 and 2.3 that

$$\begin{aligned}
l_{m,n}(t_0,\theta) &= 2\sum_{i=1}^{m+n}\gamma_i - \sum_{i=1}^{m+n}\gamma_i^2 + 2\sum_{i=1}^{m+n}\eta_i \\
&= 2(m+n)\lambda\frac{1}{m+n}\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta) - (m+n)S_{m+n}\lambda^2 + 2\sum_{i=1}^{m+n}\eta_i \\
&= \frac{(m+n)\{\frac{1}{m+n}\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta)\}^2}{S_{m+n}} - (m+n)S_{m+n}\beta_n^2 + 2\sum_{i=1}^{m+n}\eta_i \\
&= \frac{(m+n)\{\frac{1}{m+n}\sum_{i=1}^{m+n}(\hat{V}_i(t_0)-\theta)\}^2}{S_{m+n}} + o_p(1) \\
&\xrightarrow{d} \chi_1^2,
\end{aligned}$$

i.e., Theorem 2.2 holds. □

Table 2.1: Coverage probabilities for the ROC curve $R(0.05)$ are reported for the intervals based on the naive bootstrap method for $R_{m,n}(t)$ (NBM), the proposed jackknife empirical likelihood method (JELM) and the empirical likelihood method (ELM) in Claeskens et al. [11] for levels $\gamma = 0.9, 0.95$ and various sample sizes.

$(m, n, Case)$	NBM γ=0.9	JELM γ=0.9	ELM γ=0.9	NBM γ=0.95	JELM γ=0.95	ELM γ=0.95
(50, 50, A)	0.5383	0.8530	0.6816	0.5510	0.8867	0.7042
(50, 100, A)	0.5545	0.8356	0.6786	0.5692	0.8838	0.7119
(50, 200, A)	0.5324	0.8183	0.6442	0.5424	0.8708	0.6855
(100, 50, A)	0.7517	0.8950	0.8157	0.7667	0.9314	0.8488
(100, 100, A)	0.7858	0.8903	0.8329	0.8015	0.9311	0.8706
(100, 200, A)	0.7763	0.8719	0.8058	0.7880	0.9236	0.8509
(200, 50, A)	0.7331	0.9070	0.8998	0.7489	0.9473	0.9302
(200, 100, A)	0.8006	0.9147	0.9185	0.8133	0.9552	0.9495
(200, 200, A)	0.7992	0.9050	0.9144	0.8102	0.9496	0.9493
(50, 50, B)	0.1631	0.9138	0.9284	0.1645	0.9547	0.9758
(50, 100, B)	0.1431	0.8326	0.9404	0.1439	0.9351	0.9877
(50, 200, B)	0.1040	0.6433	0.9520	0.1044	0.8293	0.9897
(100, 50, B)	0.2456	0.9377	0.9544	0.2498	0.9636	0.9678
(100, 100, B)	0.2490	0.8952	0.9695	0.2522	0.9623	0.9786
(100, 200, B)	0.1962	0.7255	0.9800	0.1970	0.8845	0.9873
(200, 50, B)	0.3531	0.9448	0.9236	0.3611	0.9647	0.9288
(200, 100, B)	0.3699	0.9364	0.9415	0.3781	0.9759	0.9477
(200, 200, B)	0.3211	0.8203	0.9626	0.3248	0.9374	0.9669
(50, 50, C)	0.6505	0.9056	0.8363	0.6727	0.9550	0.8570
(50, 100, C)	0.7041	0.8686	0.8897	0.7262	0.9379	0.9149
(50, 200, C)	0.7010	0.8223	0.8944	0.7187	0.9052	0.9269
(100, 50, C)	0.7359	0.9151	0.8033	0.7572	0.9589	0.8330
(100, 100, C)	0.8208	0.9058	0.8797	0.8424	0.9532	0.9135
(100, 200, C)	0.8433	0.8656	0.9141	0.8601	0.9350	0.9507
(200, 50, C)	0.7518	0.9078	0.7349	0.7916	0.9473	0.8055
(200, 100, C)	0.8244	0.9135	0.8184	0.8681	0.9585	0.8845
(200, 200, C)	0.8562	0.8973	0.8950	0.8940	0.9508	0.9409

Table 2.2: Coverage probabilities for the ROC curve $R(0.1)$ are reported for intervals based on the naive bootstrap method for $R_{m,n}(t)$ (NBM), the proposed jackknife empirical likelihood method (JELM) and the empirical likelihood method (ELM) in Claeskens et al. [11] for levels $\gamma = 0.9, 0.95$ and various sample sizes.

$(m, n, Case)$	NBM γ=0.9	JELM γ=0.9	ELM γ=0.9	NBM γ=0.95	JELM γ=0.95	ELM γ=0.95
(50, 50, A)	0.7673	0.8685	0.8292	0.7797	0.9001	0.8664
(50, 100, A)	0.7659	0.8601	0.8101	0.7734	0.9013	0.8557
(50, 200, A)	0.7497	0.8469	0.7772	0.7561	0.8928	0.8237
(100, 50, A)	0.7478	0.8997	0.9066	0.7768	0.9364	0.9423
(100, 100, A)	0.7559	0.8991	0.9065	0.7773	0.9412	0.9411
(100, 200, A)	0.7526	0.8961	0.8955	0.7727	0.9396	0.9345
(200, 50, A)	0.8150	0.8910	0.8976	0.8739	0.937	0.9516
(200, 100, A)	0.8347	0.9040	0.9060	0.8936	0.9496	0.9594
(200, 200, A)	0.8369	0.9019	0.9019	0.9032	0.9478	0.9548
(50, 50, B)	0.4936	0.9015	0.5875	0.5121	0.9449	0.6052
(50, 100, B)	0.4539	0.8672	0.6000	0.4661	0.9348	0.6121
(50, 200, B)	0.4429	0.7871	0.6065	0.4508	0.8946	0.6206
(100, 50, B)	0.6660	0.9173	0.7102	0.6809	0.9511	0.7273
(100, 100, B)	0.6670	0.9122	0.7466	0.6758	0.9574	0.7637
(100, 200, B)	0.6615	0.8443	0.7616	0.6690	0.9302	0.7805
(200, 50, B)	0.6190	0.9140	0.7846	0.6401	0.9453	0.8116
(200, 100, B)	0.6191	0.9215	0.8356	0.6353	0.9596	0.8643
(200, 200, B)	0.6195	0.8947	0.8769	0.6319	0.9544	0.9039
(50, 50, C)	0.8103	0.9068	0.8784	0.8339	0.9524	0.9232
(50, 100, C)	0.8257	0.9078	0.9114	0.8540	0.9553	0.9502
(50, 200, C)	0.8472	0.9040	0.9094	0.8731	0.9573	0.9529
(100, 50, C)	0.7946	0.8851	0.8168	0.8521	0.9354	0.8856
(100, 100, C)	0.8360	0.9060	0.8841	0.8900	0.9530	0.9397
(100, 200, C)	0.8531	0.9068	0.9069	0.9033	0.9575	0.9516
(200, 50, C)	0.7717	0.8771	0.7655	0.8342	0.9185	0.8212
(200, 100, C)	0.8026	0.8926	0.8422	0.8668	0.9375	0.8949
(200, 200, C)	0.8274	0.9005	0.8886	0.8880	0.9512	0.9369

Table 2.3: Coverage probabilities for the ROC curve $R(0.25)$ are reported for intervals based on the naive bootstrap method for $R_{m,n}(t)$ (NBM), the proposed jackknife empirical likelihood method (JELM) and the empirical likelihood method (ELM) in Claeskens et al. [11] for levels $\gamma = 0.9, 0.95$ and various sample sizes.

$(m, n, Case)$	NBM γ=0.9	JELM γ=0.9	ELM γ=0.9	NBM γ=0.95	JELM γ=0.95	ELM γ=0.95
(50, 50, A)	0.8320	0.9047	0.9172	0.8503	0.9417	0.9587
(50, 100, A)	0.8424	0.8984	0.9070	0.8588	0.9398	0.9479
(50, 200, A)	0.8464	0.9001	0.9069	0.8604	0.9402	0.9434
(100, 50, A)	0.8369	0.8878	0.9018	0.8662	0.9407	0.9518
(100, 100, A)	0.8657	0.9013	0.9039	0.8957	0.9481	0.9516
(100, 200, A)	0.8760	0.9041	0.9008	0.9028	0.9512	0.9501
(200, 50, A)	0.8305	0.8820	0.9032	0.8786	0.9348	0.9508
(200, 100, A)	0.8577	0.8963	0.9045	0.9045	0.9453	0.9517
(200, 200, A)	0.8628	0.8980	0.9003	0.9137	0.9505	0.9507
(50, 50, B)	0.6957	0.8895	0.9002	0.7142	0.9354	0.9568
(50, 100, B)	0.7424	0.9022	0.9089	0.7601	0.9473	0.9582
(50, 200, B)	0.7647	0.9087	0.9002	0.7804	0.9509	0.9545
(100, 50, B)	0.7505	0.8739	0.8924	0.7862	0.9285	0.9399
(100, 100, B)	0.8129	0.8982	0.9085	0.8578	0.9473	0.9558
(100, 200, B)	0.8269	0.9056	0.9067	0.8782	0.9512	0.9539
(200, 50, B)	0.7512	0.8526	0.8791	0.8014	0.9057	0.9265
(200, 100, B)	0.8115	0.8794	0.9018	0.8576	0.9287	0.9449
(200, 200, B)	0.8438	0.9007	0.9098	0.8927	0.9465	0.9537
(50, 50, C)	0.8040	0.8878	0.8970	0.8599	0.9368	0.9434
(50, 100, C)	0.8417	0.9006	0.9060	0.8907	0.9465	0.9515
(50, 200, C)	0.8576	0.9083	0.9035	0.9108	0.9553	0.9537
(100, 50, C)	0.8137	0.8705	0.8785	0.8651	0.9260	0.9239
(100, 100, C)	0.8549	0.8915	0.9049	0.9109	0.9462	0.9507
(100, 200, C)	0.8708	0.9019	0.9083	0.9286	0.9507	0.9531
(200, 50, C)	0.7992	0.8638	0.8729	0.8554	0.9154	0.9197
(200, 100, C)	0.8371	0.8837	0.8957	0.8949	0.9339	0.9413
(200, 200, C)	0.8540	0.8916	0.9029	0.9155	0.9414	0.9496

Table 2.4: Interval lengths are reported for the ROC curve $R(t)$ based on the naive bootstrap method for $R_{m,n}(t)$ (NBM) and the proposed jackknife empirical likelihood method (JELM) for levels $\gamma = 0.9, 0.95$ and various sample sizes.

$(m, n, Case)$	NBM γ=0.9 $t = 0.1$	JELM γ=0.9 $t = 0.1$	NBM γ=0.95 $t = 0.1$	JELM γ=0.95 $t = 0.1$	NBM γ=0.9 $t = 0.25$	JELM γ=0.9 $t = 0.25$	NBM γ=0.95 $t = 0.25$	JELM γ=0.95 $t = 0.25$
(50, 50, A)	0.0879	0.0582	0.1031	0.0818	0.1346	0.1128	0.1590	0.1567
(50, 100, A)	0.0746	0.0573	0.0873	0.0779	0.1271	0.1098	0.1500	0.1532
(50, 200, A)	0.0700	0.0579	0.0814	0.0780	0.1205	0.1072	0.1420	0.1532
(100, 50, A)	0.0711	0.0448	0.0844	0.0653	0.1089	0.0923	0.1294	0.1334
(100, 100, A)	0.0623	0.0466	0.0736	0.0646	0.0975	0.0848	0.1158	0.1296
(100, 200, A)	0.0571	0.0447	0.0672	0.0634	0.0910	0.0804	0.1080	0.1285
(200, 50, A)	0.0599	0.0387	0.0710	0.0568	0.0883	0.0776	0.1051	0.1190
(200, 100, A)	0.0495	0.0374	0.0589	0.0539	0.0765	0.0672	0.0908	0.1135
(200, 200, A)	0.0441	0.0344	0.0524	0.0525	0.0681	0.0604	0.0811	0.1102
(50, 50, B)	0.0766	0.0415	0.0948	0.0674	0.1767	0.1284	0.2071	0.1791
(50, 100, B)	0.0540	0.0427	0.0662	0.0624	0.1572	0.1221	0.1851	0.1706
(50, 200, B)	0.0439	0.0449	0.0533	0.0610	0.1436	0.1142	0.1691	0.1682
(100, 50, B)	0.0702	0.0320	0.0855	0.0536	0.1519	0.1134	0.1793	0.1624
(100, 100, B)	0.0495	0.0317	0.0601	0.0482	0.1296	0.1038	0.1534	0.1517
(100, 200, B)	0.0395	0.0335	0.0471	0.0461	0.1124	0.0914	0.1329	0.1463
(200, 50, B)	0.0637	0.0268	0.0772	0.0453	0.1355	0.1035	0.1597	0.1527
(200, 100, B)	0.0444	0.0247	0.0535	0.0393	0.1111	0.0930	0.1316	0.1407
(200, 200, B)	0.0340	0.0255	0.0407	0.0359	0.0912	0.0764	0.1084	0.1320
(50, 50, C)	0.2139	0.1381	0.2519	0.1969	0.2873	0.2363	0.3399	0.3894
(50, 100, C)	0.1804	0.1259	0.2137	0.1903	0.2545	0.2065	0.3018	0.3789
(50, 200, C)	0.1583	0.1132	0.1870	0.1830	0.2290	0.1879	0.2711	0.3665
(100, 50, C)	0.1863	0.1245	0.2210	0.1805	0.2488	0.2137	0.2958	0.3739
(100, 100, C)	0.1480	0.1065	0.1755	0.1670	0.2059	0.1741	0.2448	0.3521
(100, 200, C)	0.1276	0.0916	0.1515	0.1622	0.1793	0.1515	0.2129	0.3439
(200, 50, C)	0.1686	0.1138	0.1982	0.1677	0.2245	0.2011	0.2660	0.3615
(200, 100, C)	0.1294	0.0977	0.1537	0.1569	0.1765	0.1550	0.2101	0.3403
(200, 200, C)	0.1026	0.0776	0.1221	0.1469	0.1453	0.1269	0.1728	0.3246

CHAPTER III

CONDITIONAL VALUE-AT-RISK IN ARCH/GARCH MODELS

Value-at-Risk is a simple and commonly used measure in risk management. When some volatility model is employed, conditional Value-at-Risk is of importance. In this and next chapter, we propose empirical likelihood methods to obtain an interval estimation for the conditional Value-at-Risk with the volatility model being an ARCH/GARCH model and a nonparametric regression, which are widely used in financial risk management. The content of this chapter is based on Y. Gong, Z. Li and L. Peng (2009), Empirical likelihood intervals for conditional Value-at-Risk in ARCH/GARCH models, *Journal of Time Series Analysis, 31, 65–75.*

3.1 Introduction

Since Engle (1982) and Bollerslev (1986) introduced the autoregressive conditional heteroscedastic (ARCH) and the generalized autoregressive conditional heteroscedastic (GARCH) models, there has been an extensive study on statistical inference, applications and extensions of ARCH/GARCH models. An interesting application field is risk management, where ARCH/GARCH models are used to model financial market volatilities; see Chapter 4 of McNeil, Frey and Embrechts (2005).

In order to assess and control the huge loss of a financial position in financial market, Value-at-Risk (VaR) becomes a widely used measure. For some given portfolio, probability and time horizon, VaR is defined as a threshold value such that the probability that the mark-to-market loss on the portfolio over the given time horizon exceeds this value (assuming normal markets and no trading) is the given probability level; see Jorion (2006). There are several approaches to compute VaR, for example, RiskMetrics developed by J.P. Morgan (1996), time series econometrics model, quantile estimation, extreme value theory, etc. We refer to Tsay (2002) for more details. In this chapter, we consider the calculation of risks combined with volatility models. In this case, a key risk measure based on the VaR concept is the conditional VaR, which is the worst possible loss at a given confidence level due to adverse market movements over the next reporting period conditional on current portfolio volatility and market information.

To make the definitions more accurate, we first introduce the ARCH/GARCH model. A GARCH(p,q) model is defined as

$$X_t = \epsilon_t \sqrt{h_t} \quad \text{and} \quad h_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{3.1}$$

where $\omega > 0, \alpha_i, \beta_j \geq 0$ are constants and ϵ_t are i.i.d. random variables with mean zero and variance one. Throughout, we denote $\lambda = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^T \in \Theta$, where Θ is a compact set of R^{p+q+1} and let $\lambda = \lambda_0$ be the true value and an interior point of Θ. If $\beta_j = 0$ for $j = 1, \ldots, q$, then (3.1) becomes an ARCH(p) model.

Suppose our observations $X_1, \cdots, X_n$ follow from the model (3.1). For $r \in (0, 1)$, the one-step ahead $100r\%$ conditional VaR, given $X_1, \cdots, X_n$, is defined as

$$q_r = \inf\{x : P(X_{n+1} \leq x | X_{n+1-k}, k \geq 1) \geq r\}.$$

An obvious estimator for the conditional VaR q_r is

$$\hat{q}_r = \sqrt{\hat{h}_{n+1}} \hat{\theta}_{\epsilon,r},$$

where $\hat{h}_t$ is an estimated conditional variance and $\hat{\theta}_{\epsilon,r}$ is an estimator of the $100r\%$ quantile of ϵ_t. To get $\hat{h}_t$, one can simply replace the parameters λ in the GARCH model by corresponding estimators, for example, the quasi-maximum likelihood estimator(QMLE); see Hall and Yao (2003) for a detailed study on the QMLE for GARCH models. Using $\hat{h}_t$, one can estimate the error ϵ_t by $\hat{\epsilon}_t = X_t / \sqrt{\hat{h}_t}$. Therefore,

$\hat{\theta}_{\epsilon,r}$ can be defined as the rth sample quantile based on the estimated errors $\hat{\epsilon}_1, \cdots, \hat{\epsilon}_n$. Recently, Tanial and Taniguchi (2008) studied the probability of $P(X_{n+1} \leq \hat{q}_r)$ under the setup of ARCH(p) models. Since the defined conditional VaR q_r is a function of $X_1, \cdots, X_n$, it is hard to evaluate the asymptotic behavior of $\hat{q}_r/q_r - 1$ unconditionally. However, when conditional high quantile is concerned, i.e., $(r = r(n) \to 1)$, Chan et al. (2007) derived the asymptotic limit of $\hat{q}_r/q_r - 1$ and thus interval estimation for q_r was obtained. In this paper, we consider the interval estimation for the conditional VaR q_r given $X_1 = z_1, \cdots, X_n = z_n$, i.e., replacing $X_1, \cdots, X_n$ in h_{n+1} by $z_1, \cdots, z_n$. In this case, the conditional VaR is not a random variable. Let's write this conditional VaR and its estimator as $q_r(z_1, \cdots, z_n)$ and $\hat{q}_r(z_1, \cdots, z_n)$, respectively. When ARCH(p) model is used, the conditional VaR only depends on the recent p observations $X_{n-p+1} = z_1, \cdots, X_n = z_p$.

It is easy to notice that the randomness in $\hat{q}_r(z_1, \cdots, z_n)$ comes from the estimators for λ and the $100r\%$ quantile of ϵ_t, and thus deriving the asymptotic limit for $\hat{q}_r(z_1, \cdots, z_n)/q_r(z_1, \cdots, z_n) - 1$ is feasible. Based on the asymptotic limit, an interval can be obtained via estimating the asymptotic variance. However, the asymptotic variance is a bit complicated. Instead of the normal approximation method, we propose to employ the empirical likelihood method to construct confidence intervals for both the quantiles of the error and the conditional VaR.

3.2 Main results

To ensure that the process $\{X_t, t = 0, \pm1, \pm2, \ldots,\}$, defined by equation (3.1), is strictly stationary with $EX_t^2 < \infty$, we need the following assumption.

Assumption 3.1. *For each* $\lambda \in \Theta$, $\omega > 0$, α_i, $\beta_i \geq 0$, $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$.

As Chan and Ling (2006) constructed empirical likelihood confidence intervals for the parameters λ, we employ the estimating equation approach in Qin and Lawless (1994) to score equations and some constraints.

By temporarily assuming that ϵ_t has Gaussian distribution, the log likelihood function can be written as

$$L(\lambda) = \sum_{t=1}^{n} l_t(\lambda) \quad \text{and} \quad l_t(\lambda) = -\frac{1}{2}\log h_t(\lambda) - \frac{X_t^2}{2h_t(\lambda)}, \tag{3.2}$$

where λ is defined in (3.1) and

$$\begin{aligned} h_t(\lambda) \ &= \tfrac{\omega}{1-\sum_{j=1}^{q}\beta_j} + \sum_{i=1}^{\min(p,t-1)} \alpha_i X_{t-i}^2 + \sum_{i=1}^{p} \alpha_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} \\ &\quad \times X_{t-i-j_1-\cdots-j_k}^2 I(t-i-j_1-\cdots-j_k \geq 1) \end{aligned}$$

is given in (2.15) of Hall and Yao (2003).

Maximizing the above log-likelihood function results in the so-called quasi maximum likelihood estimation. Some details can be found in Gouriéroux (1997) and Hall and Yao (2003). Therefore, the corresponding score equations are $\sum_{t=1}^{n} D_t(\lambda) = 0$, where $D_t(\lambda) = \frac{\partial l_t(\lambda)}{\partial \lambda} = \frac{1}{2h_t(\lambda)} \frac{\partial h_t(\lambda)}{\partial \lambda} \left\{ \frac{X_t^2}{h_t(\lambda)} - 1 \right\}$.

3.2.1 Empirical likelihood method for the quantile of ϵ_t

Estimating the properties of errors in ARCH/GARCH models has been studied in the literature; see Berkes and Horváth (2001, 2003), Berkes, Horváth and Kokoszka (2003), Horváth and Kokoszka (2001) and Horváth, Kokoszka and Teyssière (2001). For constructing a confidence interval for the quantile of the error distribution, an obvious way is the normal approximation method based on the sample quantile of the estimated errors $\hat{\epsilon}_1, \cdots, \hat{\epsilon}_n$ given in the introduction. However, the asymptotic variance is complicated since the variability of parameter estimation involves. Although empirical likelihood method for quantiles has been proposed by Chen and Hall (1993), a straight application to the estimated errors does not lead to a chi-square limit because of the effect of parameter estimation for λ. Here we propose the following estimating equation approach to take the variability of parameter estimation into account.

Let K be a symmetric density function with support in $[-1, 1]$ and have continuous first derivative. Put $G(x) = \int_{-\infty}^{x} K(y)dy$. Then, we can construct the empirical likelihood ratio as

$$\begin{aligned} & L_n(r;\theta,\lambda) \\ = \quad & \sup\{\prod_{t=1}^{n}(np_t) : \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t G(\frac{\theta-\epsilon_t(\lambda)}{h}) = r, \sum_{t=1}^{n} p_t D_t(\lambda) = 0, p_i > 0, i = 1,\dots,n\}, \end{aligned}$$

where $\epsilon_t(\lambda) = \frac{X_t}{\sqrt{h_t(\lambda)}}$, $r \in (0,1)$ and $h = h(n) > 0$ is a bandwidth. Denote

$$g_t(\theta,\lambda) = (G(\frac{\theta-\epsilon_t(\lambda)}{h}) - r, D_t^T(\lambda))^T = (\omega_t(\lambda), D_t^T(\lambda))^T.$$

Then, the empirical likelihood ratio can be rewritten as

$$L_n(r;\theta,\lambda) = \sup\{\prod_{t=1}^{n}(np_t) : \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t g_t(\theta,\lambda) = 0, p_i > 0, i = 1,\dots,n\}. \quad (3.3)$$

Using the method of Lagrange multipliers, we find that the maximum is achieved at

$$p_t = \frac{1}{n\{1 + b^T(\theta,\lambda)g_t(\theta,\lambda)\}}, \quad t = 1,\dots,n,$$

which gives the log empirical likelihood ratio

$$l_n(r;\theta,\lambda) = 2\sum_{t=1}^{n} \log\left\{1 + b^T(\theta,\lambda)g_t(\theta,\lambda)\right\},$$

where $b(\theta,\lambda)$ satisfies

$$\frac{1}{n}\sum_{t=1}^{n} \frac{g_t(\theta,\lambda)}{1 + b^T(\theta,\lambda)g_t(\theta,\lambda)} = 0.$$

Since we are interested in θ, we consider the profiled likelihood ratio $l_n(r;\theta,\hat{\lambda}(\theta))$, where $\hat{\lambda}(\theta) = \arg\min_{\lambda} l_n(r;\theta,\lambda)$. Our main results are as follows. Here we use θ_0 to denote the true $100r\%$ quantile of ϵ_t.

Proposition 3.1. *(Gong, Li and Peng, 2009) Suppose Assumption 3.1 holds and* $E|\epsilon_t|^{4+\delta} < \infty$ *for some* $\delta > 0$. *Further assume the density of* ϵ_t *is positive at* θ_0, *its first derivative is continuous at* θ_0, *and* $n^{1-\sigma}h^2 \to \infty$ *and* $nh^4 \to 0$ *for some* $\sigma \in (0,1/2)$ *as* $n \to \infty$. *Then, with probability tending to one,* $l_n(r;\theta_0,\lambda)$ *attains its minimum value at some point* $\hat{\lambda}_n(\theta_0)$ *in the interior of* $V_n = \{\lambda : |\lambda - \lambda_0| \le n^{-0.5+\sigma/4}\}$.

Theorem 3.1. *(Gong, Li and Peng, 2009) Under conditions of Proposition 3.1, we have* $l_n(r;\theta_0,\hat{\lambda}_n(\theta_0)) \xrightarrow{d} \chi_1^2$ *as* $n \to \infty$*, where* $\hat{\lambda}_n(\theta_0)$ *is given in Proposition 3.1.*

Based on the above theorem, a confidence interval of θ_0 with level γ can be obtained as

$$I_\gamma(r) = \{\theta : l_n(r;\theta,\hat{\lambda}_n(\theta)) \le \chi^2_{1,\gamma}\},$$

where $\chi^2_{1,\gamma}$ is the γth quantile of χ_1^2.

3.2.2 Empirical likelihood confidence interval for the conditional VaR

In this subsection, we focus on an ARCH(p) model, i.e., assume $\beta_j = 0$, for $j = 1,\dots,q$ in (3.1). In this case, we only need to condition on the recent p observations as mentioned in the introduction. That is, we want to construct an empirical likelihood confidence interval for the conditional VaR

$$q_r = \inf\{x : P(X_{n+1} \le x | X_{n+1-i} = z_{p+1-i}, 1 \le i \le p) \ge r\}.$$

We employ a similar procedure as in the above subsection. Define

$$g_t(\theta,\lambda) = (G(\frac{\theta/\sqrt{h_{n+1}(\lambda)} - \epsilon_t(\lambda)}{h}) - r, D_t^T(\lambda))^T = (\omega_t(\lambda), D_t^T(\lambda))^T, \tag{3.4}$$

where θ denotes the conditional VaR given $X_n = z_p, \cdots, X_{n+1-p} = z_1$, and $h_{n+1}(\lambda) = \omega + \sum_{i=1}^p \alpha_i z_{p+1-i}^2$. As in the above subsection, we use θ_0 to denote the true value of θ, and $l_n(r;\theta,\lambda)$ is defined as in Section 2.1 except that $g_t(\theta,\lambda)$ is replaced by (3.4).

Proposition 3.2. *(Gong, Li and Peng, 2009) Suppose Assumption 3.1 holds and* $E|\epsilon_t|^{4+\delta} < \infty$ *for some* $\delta > 0$*. Further assume the density of* ϵ_t *is positive at* $\theta_0/h_{n+1}(\lambda_0)$*, its first derivative is continuous at* $\theta_0/h_{n+1}(\lambda_0)$*, and* $n^{1-\sigma}h^2 \to \infty$ *and* $nh^4 \to 0$ *for some* $\sigma \in (0,1/2)$ *as* $n \to \infty$*. Then, with probability tending to one,* $l_n(r;\theta_0,\lambda)$ *attains its minimum value at some point* $\hat{\lambda}_n(\theta_0)$ *in the interior of* $V_n = \{\lambda : |\lambda - \lambda_0| \le n^{-0.5+\sigma/4}\}$.

Theorem 3.2. *(Gong, Li and Peng, 2009) Under conditions of Proposition 3.2, we have* $l_n(r;\theta_0,\hat{\lambda}_n(\theta_0)) \xrightarrow{d} \chi_1^2$ *as* $n \to \infty$*, where* $\hat{\lambda}_n(\theta_0)$ *is given in Proposition 3.2.*

Based on the above theorem, a confidence interval of θ_0 with level γ can be obtained as

$$I_\gamma(r) = \{\theta : l_n(r;\theta,\hat{\lambda}_n(\theta)) \le \chi^2_{1,\gamma}\},$$

where $\chi^2_{1,\gamma}$ is the γth quantile of χ_1^2.

Remark 3.1. *Motivated by the optimal choice of bandwidth in smoothing distribution estimation (see Cheng and Peng (2002)), we propose to choose* $h = cn^{-1/3}$ *for some positive* c *in both Theorems 3.1 and 3.2. In the simulation study below, we use* $c = 1$*. Since we do not have the expansion for the coverage accuracy, the theoretical optimal bandwidth is not available.*

3.3 Simulations

In this section, a simulation study is carried out to assess the accuracy of the proposed empirical likelihood confidence intervals for the conditional VaR in Section 3.2. As mentioned in Section 3.2, the normal approximation based confidence intervals involve complicated asymptotic variance, we will not conduct a comparison. We choose ARCH(1) models and consider conditional VaR at $X_n = 0$ and 1 with level $r = 0.90$, 0.95, and 0.99. In each case, data are generated from an ARCH(1) model with parameters $(\omega, \alpha_1) = (0.5, 0.7)$ or $(0.5, 0.9)$, and the error ϵ_t has a $N(0,1)$ distribution and a standardized t-distribution with the degree of freedom 5, i.e., $\sqrt{\frac{3}{5}}t(5)$. From each model, we draw random samples with the sample sizes $n = 1000$ and 3000. The number of replications is set to 5000. We employ the kernel $K(x) = \frac{15}{16}(1-x^2)^2 I(|x| \le 1)$ and choose the bandwidth $h = n^{-1/3}$.

Based on the 5000 replications, we report the empirical coverage probabilities with confidence levels $\gamma = 0.90$, 0.95 in Table 3.1. From Table 3.1, we observe that the proposed method has good coverage accuracy for the cases of $r = 0.90$ and 0.95. The worst coverage accuracy is the cases when $n = 1000$ and $r = 0.99$. This is because the quantile is large with respect to the sample size. When the sample size increases from $n = 1000$ to 3000, the coverage probabilities for the case of $r = 0.99$ become much better. In most cases, sample size $n = 3000$ performs better than $n = 1000$ and the cases of having the normal distribution give much accurate coverage probability than the case of having the t(5) distribution.

3.4 Proofs

First we list some lemmas.

Lemma 3.1. *(Ling, 2007) Under Assumption 3.1, we have, for any constant $\iota > 0$,*

$$\sup_{\lambda \in V_0} \left\| \frac{1}{h_t(\lambda)} \frac{\partial h_t(\lambda)}{\partial \lambda} \right\| \leq C(1 + \sum_{i=1}^{\infty} \rho^i |X_{t-i}|)^{\iota}$$

and

$$\sup_{\lambda \in V_0} \left\| \frac{1}{h_t(\lambda)} \frac{\partial^2 h_t(\lambda)}{\partial \lambda \partial \lambda^T} \right\| \leq C(1 + \sum_{i=1}^{\infty} \rho^i |X_{t-i}|)^{\iota},$$

where V_0 is some neighborhood of λ_0, $C > 0$ and $\rho \in (0, 1)$ are constants.

Lemma 3.2. *Put $A_n(\lambda) = \frac{1}{n} \sum_{t=1}^{n} \frac{1}{h} K(\frac{\theta_0 - \epsilon_t(\lambda)}{h}) \frac{\partial \epsilon_t(\lambda)}{\partial \lambda}$. Then, under conditions of Proposition 3.1, we have*

$$\sup_{\lambda \in V_n} \| A_n(\lambda) - A_n(\lambda_0) \| = o_p(1), \tag{3.5}$$

where V_n is defined in Proposition 3.1.

Proof. Since

$$\frac{\partial \epsilon_t(\lambda)}{\partial \lambda} = -\epsilon_t(\lambda) \frac{1}{2h_t(\lambda)} \frac{\partial h_t(\lambda)}{\partial \lambda} \tag{3.6}$$

and

$$\frac{\partial^2 \epsilon_t(\lambda)}{\partial \lambda \partial \lambda^T} = \frac{3}{4} \epsilon_t(\lambda) \frac{1}{h_t^2(\lambda)} \frac{\partial h_t(\lambda)}{\partial \lambda} \frac{\partial h_t(\lambda)}{\partial \lambda^T} - \frac{1}{2} \epsilon_t(\lambda) \frac{1}{h_t(\lambda)} \frac{\partial^2 h_t(\lambda)}{\partial \lambda \partial \lambda^T}, \tag{3.7}$$

we have

$$\begin{aligned}\|A_n(\lambda)-A_n(\lambda_0)\| &\leq \frac{1}{nh}\sum_{t=1}^{n}\left\|K\left(\frac{\theta_0-\epsilon_t(\lambda)}{h}\right)\left(\frac{\partial\epsilon_t(\lambda)}{\partial\lambda}-\frac{\partial\epsilon_t(\lambda_0)}{\partial\lambda}\right)\right\| \\ &\quad +\frac{1}{nh}\sum_{t=1}^{n}\left|K\left(\frac{\theta_0-\epsilon_t(\lambda)}{h}\right)-K\left(\frac{\theta_0-\epsilon_t(\lambda_0)}{h}\right)\right|\left\|\frac{\partial\epsilon_t(\lambda_0)}{\partial\lambda}\right\| \\ &= B_1+B_2.\end{aligned}$$

Denote $\xi_{t-1}=C(1+\sum_{i=1}^{\infty}\rho^i|X_{t-i}|)^{\sigma/4}$. Then $E\xi_{t-1}^{\frac{8}{\sigma}}<\infty$, which implies

$$\max_{1\leq t\leq n}\xi_{t-1}=O_p(n^{\sigma/8}). \tag{3.8}$$

By the mean value theorem, (3.7), (3.8), Lemma 3.1 and the condition on h, we have

$$\begin{aligned} & B_1 \\ \leq\ & \tfrac{1}{nh}\textstyle\sum_{t=1}^{n}K\left(\frac{\theta_0-\epsilon_t(\lambda)}{h}\right)\left\|\frac{\partial^2\epsilon_t(\lambda^*)}{\partial\lambda\partial\lambda^T}\right\|\|\lambda-\lambda_0\| \\ \leq\ & \tfrac{1}{nh}\textstyle\sum_{t=1}^{n}|\epsilon_t(\lambda^*)|K\left(\frac{\theta_0-\epsilon_t(\lambda)}{h}\right)\left\{\frac{3}{4}\left\|\frac{1}{h_t(\lambda^*)}\frac{\partial h_t(\lambda^*)}{\partial\lambda}\right\|^2+\frac{1}{2}\left\|\frac{1}{h_t(\lambda^*)}\frac{\partial^2 h_t(\lambda^*)}{\partial\lambda\partial\lambda^T}\right\|\right\}\|\lambda-\lambda_0\| \\ \leq\ & \sup_x|K(x)|\left\{\tfrac{2}{\sqrt{\omega_0}}\tfrac{1}{n}\textstyle\sum_{t=1}^{n}|X_t|\right\}\tfrac{\|\lambda-\lambda_0\|}{h}\{\tfrac{3}{4}\max_{1\leq t\leq n}\xi_{t-1}^2+\tfrac{1}{2}\max_{1\leq t\leq n}\xi_{t-1}\} \\ \xrightarrow{p}\ & 0 \end{aligned} \tag{3.9}$$

uniformly in $\lambda\in V_n$, where each element of λ^* lies between the corresponding ones of λ_0 and λ. Similarly, by (3.6), we get

$$\begin{aligned} & B_2 \\ \leq\ & \tfrac{1}{nh^2}\textstyle\sum_{t=1}^{n}\left|K'\left(\frac{\theta_0-\epsilon_t(\lambda^*)}{h}\right)\right|\left\|\frac{\partial\epsilon_t(\lambda^*)}{\partial\lambda}\right\|\left\|\frac{\partial\epsilon_t(\lambda_0)}{\partial\lambda}\right\|\|\lambda-\lambda_0\| \\ \leq\ & \tfrac{1}{2}\left\{\tfrac{1}{nh}\textstyle\sum_{t=1}^{n}\left|K'\left(\frac{\theta_0-\epsilon_t(\lambda^*)}{h}\right)\right||\epsilon_t(\lambda^*)\epsilon_t(\lambda_0)|\right\}\tfrac{\|\lambda-\lambda_0\|}{h}\max_{1\leq t\leq n}\xi_{t-1}^2. \end{aligned} \tag{3.10}$$

Now,

$$\begin{aligned} & \tfrac{1}{nh}\textstyle\sum_{t=1}^{n}|K'(\frac{\theta_0-\epsilon_t(\lambda^*)}{h})||\epsilon_t(\lambda^*)\epsilon_t(\lambda_0)| \\ \leq\ & \tfrac{1}{nh}\textstyle\sum_{t=1}^{n}|K'(\frac{\theta_0-\epsilon_t(\lambda^*)}{h})||\epsilon_t^2(\lambda^*)-\theta_0^2|+\tfrac{1}{nh}\textstyle\sum_{t=1}^{n}|K'(\frac{\theta_0-\epsilon_t(\lambda^*)}{h})|\theta_0^2 \\ & +\tfrac{1}{nh}\textstyle\sum_{t=1}^{n}|K'(\frac{\theta_0-\epsilon_t(\lambda^*)}{h})||\epsilon_t(\lambda^*)\epsilon_t(\lambda_0)-\epsilon_t^2(\lambda^*)| \\ =\ & D_1+D_2+D_3. \end{aligned}$$

By the ergodic theorem, we get

$$\begin{aligned}
& D_1 \\
= \; & \tfrac{1}{n}\textstyle\sum_{t=1}^n |K'(\frac{\theta_0-\epsilon_t(\lambda^*)}{h})||\frac{\epsilon_t(\lambda^*)-\theta_0}{h}||\epsilon_t(\lambda^*)+\theta_0| \\
\leq \; & \sup_x |K'(x)|\tfrac{1}{n}\textstyle\sum_{t=1}^n |\epsilon_t(\lambda^*)| + |\theta_0|\sup_x|K'(x)| \\
\leq \; & \sup_x |K'(x)|\tfrac{2}{\sqrt{\omega_0}}\tfrac{1}{n}\textstyle\sum_{t=1}^n |X_t| + |\theta_0|\sup_x|K'(x)| \\
\xrightarrow{p} \; & \sup_x |K'(x)|\tfrac{2}{\sqrt{\omega_0}}E|X_t| + |\theta_0|\sup_x|K'(x)|
\end{aligned}$$

and

$$\begin{aligned}
& D_3 \\
\leq \; & \tfrac{1}{2nh}\textstyle\sum_{t=1}^n |K'(\frac{\theta_0-\epsilon_t(\lambda^*)}{h})||\epsilon_t(\lambda^*)\epsilon_t(\lambda^{**})|\|\lambda^*-\lambda_0\|\xi_{t-1} \\
\leq \; & \sup_x |K'(x)|\tfrac{1}{\omega_0}\left\{\tfrac{1}{n}\textstyle\sum_{t=1}^n X_t^2\right\}\frac{\|\lambda^*-\lambda_0\|}{h}\max_{1\leq t\leq n}\xi_{t-1} \\
\xrightarrow{p} \; & 0,
\end{aligned}$$

where each element of λ^{**} lies between the corresponding ones of λ^* and λ_0. It follows from (3.8) that

$$\begin{aligned}
& |\epsilon_t(\lambda)-\epsilon_t(\lambda_0)| \\
\leq \; & \|\tfrac{\partial\epsilon_t(\lambda^*)}{\partial\lambda}\|\|\lambda-\lambda_0\| \\
\leq \; & \tfrac{1}{2}|\epsilon_t(\lambda^*)|\xi_{t-1}\|\lambda-\lambda_0\| \\
\leq \; & \tfrac{1}{2}|\epsilon_t(\lambda_0)|\xi_{t-1}\|\lambda-\lambda_0\| + \tfrac{1}{4}|\epsilon_t(\lambda^{**})|\xi_{t-1}^2\|\lambda-\lambda_0\|\|\lambda^*-\lambda_0\| \\
\leq \; & \tfrac{1}{2}\max_{1\leq t\leq n}|\epsilon_t(\lambda_0)|\max_{1\leq t\leq n}\xi_{t-1}\|\lambda-\lambda_0\| \\
& + \tfrac{1}{2\sqrt{\omega_0}}\max_{1\leq t\leq n}|X_t|\max_{1\leq t\leq n}\xi_{t-1}^2\|\lambda-\lambda_0\|\|\lambda^*-\lambda_0\| \\
= \; & o_p(1)
\end{aligned}$$

uniformly in $t = 1, \cdots, n$ and $\lambda \in V_n$, where each element of λ^* lies between the corresponding ones of λ_0 and λ, and each element of λ^{**} lies between the corresponding ones of λ_0 and λ^*. Hence, it follows from Theorem A of Silverman (1978) that $\sup_{\lambda^*\in V_n} D_2 = O_p(1)$. Therefore, by (3.8) and the condition on h, we obtain

$$B_2 \xrightarrow{p} 0$$

uniformly in $\lambda \in V_n$. Hence Lemma 3.2 holds. □

Lemma 3.3. *(Chan and Ling, 2006) Denote* $P_t(\lambda) = \partial D_t(\lambda)/\partial\lambda^T$*. Under conditions of Proposition 3.1, we have*

1. $\sup_{\lambda\in V_n} \max_{1\le t\le n} \|P_t(\lambda)\| = o_p(n^{1-\sigma/2})$,

2. $\sup_{\lambda\in V_n} \max_{1\le t\le n} \|D_t(\lambda)\| = o_p(n^{\frac{1}{2}-\sigma/4})$,

3. $\sup_{\lambda\in V_n} \|\frac{1}{n}\sum_{t=1}^n P_t(\lambda) - \frac{1}{n}\sum_{t=1}^n P_t(\lambda_0)\| = o_p(1)$,

4. $\sup_{\lambda\in V_n} \|\frac{1}{n}\sum_{t=1}^n D_t(\lambda)D_t^T(\lambda) - \frac{1}{n}\sum_{t=1}^n D_t(\lambda_0)D_t^T(\lambda_0)\| = o_p(1)$.

Proof of Proposition 3.1. Let $b = \rho u$ with $\|u\| = 1$. Define

$$g_t(\lambda) = g_t(\theta_0, \lambda) \quad \text{and} \quad Q_{1n}(\lambda, b) = \frac{1}{n}\sum_{t=1}^n \frac{g_t(\lambda)}{1 + b^T g_t(\lambda)}.$$

It is easy to check that

$$\begin{aligned} 0 &= \|Q_{1n}(\lambda, \rho u)\| \\ &\ge |u^T Q_{1n}(\lambda, \rho u)| \\ &\ge \frac{\rho u^T S_n(\lambda) u}{1 + \rho Z_n(\lambda)} - \frac{1}{n}\left|\sum_{t=1}^n u^T g_t(\lambda)\right|, \end{aligned}$$

where $Z_n(\lambda) = \max_{1\le t\le n}\|g_t(\lambda)\|$, $S_n(\lambda) = \frac{1}{n}\sum_{t=1}^n g_t(\lambda)g_t^T(\lambda)$. By Lemma 3.2, Lemma 3.3 (c) and the central limit theorem, we get

$$\begin{aligned} & \tfrac{1}{n}|\textstyle\sum_{t=1}^n u^T g_t(\lambda)| \\ \le\ & \tfrac{1}{n}|\textstyle\sum_{t=1}^n u^T g_t(\lambda_0)| + \tfrac{1}{n}\|\textstyle\sum_{t=1}^n u^T \frac{\partial g_t(\lambda^*)}{\partial\lambda^T}\|\|\lambda - \lambda_0\| \\ =\ & O_p(n^{-0.5+\sigma/4}) \end{aligned} \tag{3.11}$$

uniformly in $\lambda \in V_n$. From Lemma 3.3 (b),

$$Z_n(\lambda) = \max_{1\le t\le n}\|g_t(\lambda)\| = \max_{1\le t\le n}\|D_t(\lambda)\| + O_p(1) = o_p(n^{0.5-\sigma/4})$$

uniformly in $\lambda \in V_n$.

By Lemma 3.2,

$$
\begin{array}{rl}
 & |\frac{1}{n}\sum_{t=1}^{n}\omega_t^2(\lambda)-\frac{1}{n}\sum_{t=1}^{n}\omega_t^2(\lambda_0)| \\
\leq & \frac{4}{n}\sum_{t=1}^{n}|\omega_t(\lambda)-\omega_t(\lambda_0)| \\
\leq & \frac{4}{n}\sum_{t=1}^{n}\|\frac{\partial\omega_t(\lambda^*)}{\partial\lambda}\|\|\lambda-\lambda_0\| \\
= & o_p(1)
\end{array}
\tag{3.12}
$$

uniformly in $\lambda \in V_n$. Using Lemma 3.3(b), Lemma 3.3(c) and the second inequality in (3.12), we can show that

$$
\begin{array}{rl}
 & \|\frac{1}{n}\sum_{t=1}^{n}\omega_t(\lambda)D_t(\lambda)-\frac{1}{n}\sum_{t=1}^{n}\omega_t(\lambda_0)D_t(\lambda_0)\| \\
\leq & \frac{2}{n}\sum_{t=1}^{n}\|P_t(\lambda^*)\|\|\lambda-\lambda_0\|+\frac{1}{n}\sum_{t=1}^{n}|\omega_t(\lambda)-\omega_t(\lambda_0)|\max_{1\leq t\leq n}\|D_t(\lambda_0)\| \\
= & o_p(1)
\end{array}
\tag{3.13}
$$

uniformly in $\lambda \in V_n$. So, it follows from Lemma 3.3 (d), (3.12) and (3.13) that

$$
S_n(\lambda) = S_n(\lambda_0) + o_p(1) \xrightarrow{p} \hat{\Omega}, \tag{3.14}
$$

where $\hat{\Omega}$ is some positive definite matrix in $R^{(p+q+2)\times(p+q+2)}$. Therefore, we obtain

$$
\rho = O_p(n^{-0.5+\sigma/4}). \tag{3.15}
$$

Put $\gamma_t(\lambda) = b^T g_t(\lambda)$. Then

$$
\max_{1\leq t\leq n}|\gamma_t(\lambda)| \leq \|b^T\|\max_{1\leq t\leq n}\|g_t(\lambda)\| = o_p(1).
$$

It follows from Taylor expansion, Lemma 3(d) and (4.15) that

$$
\begin{array}{rcl}
0 = Q_{1n}(\lambda, b) & = & \frac{1}{n}\sum_{t=1}^{n}g_t(\lambda) - S_n(\lambda)b + \frac{1}{n}\sum_{t=1}^{n}\frac{\gamma_t^2 g_t(\lambda)}{1+\gamma_t} \\
 & \leq & \frac{1}{n}\sum_{t=1}^{n}g_t(\lambda) - S_n(\lambda)b + \frac{\|b\|\max_{1\leq t\leq n}|\gamma_t|}{1-\max_{1\leq t\leq n}|\gamma_t|}\frac{1}{n}\sum_{t=1}^{n}\|g_t(\lambda)\|^2 \\
 & = & \frac{1}{n}\sum_{t=1}^{n}g_t(\lambda) - S_n(\lambda)b + o_p(n^{-0.5+\sigma/4})
\end{array}
$$

uniformly in $\lambda \in V_n$, which implies that

$$
b = S_n^{-1}(\lambda)\left\{\frac{1}{n}\sum_{t=1}^{n}g_t(\lambda)\right\} + o_p(n^{-0.5+\sigma/4}) \tag{3.16}
$$

uniformly in $\lambda \in V_n$.

Next, put $\lambda = \lambda_0 + vn^{-0.5+\sigma/4}$ with $\|v\| = 1$. By (3.16), (3.14), Lemma 3.2 and Lemma 3.3 (c), we have

$$\begin{aligned}
& l_n(r;\theta_0,\lambda) \\
= \ & 2\sum_{t=1}^n b^T(\lambda)g_t(\lambda) - \sum_{t=1}^n \{b^T(\lambda)g_t(\lambda)\}^2 + o_p(n^{\sigma/2}) \\
= \ & n\left\{\tfrac{1}{n}\sum_{t=1}^n g_t(\lambda)\right\}^T S_n(\lambda)^{-1}\left\{\tfrac{1}{n}\sum_{t=1}^n g_t(\lambda)\right\} + o_p(n^{\sigma/2}) \\
= \ & n\left\{\tfrac{1}{n}\sum_{t=1}^n g_t(\lambda_0) + \tfrac{1}{n}\sum_{t=1}^n \tfrac{\partial g_t(\lambda^*)}{\partial \lambda^T} vn^{-0.5+\sigma/4}\right\}^T S_n(\lambda)^{-1} \\
& \times \left\{\tfrac{1}{n}\sum_{t=1}^n g_t(\lambda_0) + \tfrac{1}{n}\sum_{t=1}^n \tfrac{\partial g_t(\lambda^*)}{\partial \lambda^T} vn^{-0.5+\sigma/4}\right\} + o_p(n^{\sigma/2}) \\
= \ & n\left\{O_p\{n^{-1/2}(\log\log n)^{1/2}\} + E\{\tfrac{\partial g_t(\lambda_0)}{\partial \lambda^T}\}vn^{-0.5+\sigma/4}\right\}^T \hat{\Omega}^{-1} \\
& \times \left\{O_p\{n^{-1/2}(\log\log n)^{1/2}\} + E\{\tfrac{\partial g_t(\lambda_0)}{\partial \lambda^T}\}vn^{-0.5+\sigma/4}\right\} + o_p(n^{\sigma/2}) \\
\geq \ & \tfrac{c}{2}n^{\sigma/2}
\end{aligned} \tag{3.17}$$

uniformly in $\|v\| = 1$ in probability, where $c > 0$ is the smallest eigenvalue of

$$E\left\{\frac{\partial g_t(\lambda_0)}{\partial \lambda^T}\right\}^T \hat{\Omega}^{-1} E\left\{\frac{\partial g_t(\lambda_0)}{\partial \lambda^T}\right\}.$$

Also, from (3.17), we can see that

$$l_n(r;\theta_0,\lambda_0) = O_p(\log\log n) < l_n(r;\theta_0,\lambda)$$

with probability tending to one. Since $l_n(r;\theta_0,\lambda)$ is a continuous function with respect to $\lambda \in V_n$, $l_n(r;\theta_0,\lambda)$ has minimum value in some interior point $\hat{\lambda}_n(\theta_0) \in V_n$ and $\hat{\lambda}_n(\theta_0)$ satisfies

$$\begin{aligned}
& \tfrac{\partial l_n(r;\theta_0,\lambda)}{\partial \lambda} \big|_{\lambda=\hat{\lambda}_n(\theta_0)} \\
= \ & \sum_{t=1}^n \tfrac{2}{1+b^T(\lambda)g_t(\lambda)}\left\{\tfrac{\partial g_t(\lambda)}{\partial \lambda^T}\right\}^T b(\lambda) \big|_{\lambda=\hat{\lambda}_n(\theta_0)} \\
= \ & 0.
\end{aligned} \tag{3.18}$$

□

Proof of Theorem 3.1. Define

$$Q_{2n}(\lambda, b) = \frac{1}{n}\sum_{t=1}^n \frac{1}{1+b^T g_t(\lambda)}\left\{\frac{\partial g_t(\lambda)}{\partial \lambda^T}\right\}^T b.$$

It follows from (3.18) that

$$Q_{2n}(\hat{\lambda}_n, b(\hat{\lambda}_n)) = 0,$$

where, for simplicity, we put $\hat{\lambda}_n = \hat{\lambda}_n(\theta_0)$.

Since

$$\frac{\partial Q_{2n}(\lambda, b)}{\partial \lambda_i} = \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{1}{1+b^T g_t(\lambda)}\frac{\partial}{\partial \lambda_i}\left(\frac{\partial g_t(\lambda)}{\partial \lambda^T}\right)^T b - \left(\frac{\partial g_t(\lambda)}{\partial \lambda^T}\right)^T b\frac{b^T \frac{\partial g_t(\lambda)}{\partial \lambda_i}}{\{1+b^T g_t(\lambda)\}^2}\right\},$$

we have

$$\begin{aligned} & \left\|\frac{\partial Q_{2n}(\lambda, b)}{\partial \lambda_i}\right\| \\ \leq & \frac{\|b\|}{1-\max|\gamma_t|}\frac{1}{n}\textstyle\sum_{t=1}^{n}\left\|\frac{\partial}{\partial \lambda_i}(\frac{\partial g_t(\lambda)}{\partial \lambda^T})\right\| \\ & +\frac{\|b\|^2 \max\limits_{1\leq t\leq n}\|\frac{\partial g_t(\lambda)}{\partial \lambda^T}\|}{\left(1-\max\limits_{1\leq t\leq n}|\gamma_t|\right)^2}\frac{1}{n}\textstyle\sum_{t=1}^{n}\|\frac{\partial g_t(\lambda)}{\partial \lambda^T}\|, \end{aligned} \tag{3.19}$$

where γ_t is given in the proof of Proposition 3.1. By Lemma 6.1(c) of Chan and Ling (2006) and in a way similar to the proof of (3.9) and (3.10) in Lemma 3.2, we can show that the first term in (3.19) is $o_p(1)$ uniformly in $\lambda \in V_n$. By (3.6) and Lemma 3.2,

$$\begin{aligned} & \sup_{\lambda\in V_n}\|\frac{1}{h}K(\tfrac{\theta_0-\epsilon_t(\lambda)}{h})\tfrac{\partial \epsilon_t(\lambda)}{\partial \lambda^T}\| \\ \leq & \sup_{\lambda\in V_n}|\tfrac{1}{2}\tfrac{1}{h}K(\tfrac{\theta_0-\epsilon_t(\lambda)}{h})\epsilon_t(\lambda)\xi_{t-1}| \\ \leq & \sup_{\lambda\in V_n}|\tfrac{1}{2}K(\tfrac{\theta_0-\epsilon_t(\lambda)}{h})\tfrac{\epsilon_t(\lambda)-\theta_0+\theta_0}{h}|\max_{1\leq t\leq n}\xi_{t-1} \\ \leq & \tfrac{1}{2}\{\sup_x K(x)\}(1+\tfrac{\theta_0}{h})\max_{1\leq t\leq n}\xi_{t-1}. \end{aligned}$$

It follows from (3.8), (4.15), Lemma 3.2 and conditions on h that the second term is also $o_p(1)$ uniformly in $\lambda \in V_n$, which implies that

$$\frac{\partial Q_{2n}(\lambda, b)}{\partial \lambda^T} \xrightarrow{p} 0$$

uniformly in $\lambda \in V_n$.

Note that

$$\frac{\partial Q_{2n}(\lambda, b)}{\partial b^T} = \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{1}{1+b^T g_t(\lambda)}\left(\frac{\partial g_t(\lambda)}{\partial \lambda^T}\right)^T - \left(\frac{\partial g_t(\lambda)}{\partial \lambda^T}\right)^T b\frac{g_t^T(\lambda)}{\{1+b^T g_t(\lambda)\}^2}\right\}. \tag{3.20}$$

From Lemma 3.2 and Lemma 3.3(c), we know that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g_t(\lambda)}{\partial \lambda^T}\xrightarrow{p}\Omega$$

uniformly in $\lambda \in V_n$, where Ω is some matrix in $R^{(p+q+2)\times(p+q+1)}$. Hence, we obtain

$$\begin{aligned}
& \|\tfrac{\partial Q_{2n}(\lambda,b)}{\partial b^T}-\Omega^T\| \\
\le\ & \frac{\max\limits_{1\le t\le n}|\gamma_t|}{1-\max\limits_{1\le t\le n}|\gamma_t|}\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\|\tfrac{\partial g_t(\lambda)}{\partial \lambda^T}\|+\|\tfrac{1}{n}\sum_{t=1}^{n}(\tfrac{\partial g_t(\lambda)}{\partial \lambda^T})^T-\Omega^T\| \\
& +\frac{\max\limits_{1\le t\le n}|\gamma_t|}{(1-\max\limits_{1\le t\le n}|\gamma_t|)^2}\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\|\tfrac{\partial g_t(\lambda)}{\partial \lambda^T}\| \\
\xrightarrow{p}\ & 0
\end{aligned}$$

uniformly in $\lambda \in V_n$.

It follows from (3.20) that

$$\begin{aligned}
& \tfrac{\partial Q_{1n}(\lambda,b)}{\partial \lambda^T} \\
=\ & \tfrac{1}{n}\textstyle\sum_{t=1}^{n}\left\{\tfrac{1}{1+b^Tg_t(\lambda)}\tfrac{\partial g_t(\lambda)}{\partial \lambda^T}-\tfrac{g_t(\lambda)b^T\frac{\partial g_t(\lambda)}{\partial \lambda^T}}{\{1+b^Tg_t(\lambda)\}^2}\right\} \\
=\ & \left\{\tfrac{\partial Q_{2n}(\lambda,b)}{\partial b^T}\right\}^T \\
\xrightarrow{p}\ & \Omega
\end{aligned}$$

uniformly in $\lambda \in V_n$.

By (3.14),

$$\frac{\partial Q_{1n}(\lambda,b)}{\partial b^T}=-\frac{1}{n}\sum_{t=1}^{n}\frac{g_t(\lambda)g_t^T(\lambda)}{\{1+b^Tg_t(\lambda)\}^2}=-S_n(\lambda_0)+o_p(1)\xrightarrow{p}-\hat{\Omega}$$

uniformly in $\lambda \in V_n$.

Denote

$$M_n=M_n(\lambda,b)=\begin{pmatrix}\frac{\partial Q_{1n}}{\partial b^T} & \frac{\partial Q_{1n}}{\partial \lambda^T}\\ \frac{\partial Q_{2n}}{\partial b^T} & \frac{\partial Q_{2n}}{\partial \lambda^T}\end{pmatrix}.$$

Hence

$$M_n\xrightarrow{p}\begin{pmatrix}-\hat{\Omega} & \Omega\\ \Omega^T & 0\end{pmatrix}\triangleq M=\begin{pmatrix}M_{11} & M_{12}\\ M_{21} & M_{22}\end{pmatrix}$$

uniformly in $\lambda \in V_n$. Using Taylor expansions and the facts that

$$Q_{1n}(\hat{\lambda}_n, b(\hat{\lambda}_n)) = 0 \quad \text{and} \quad Q_{2n}(\hat{\lambda}_n, b(\hat{\lambda}_n)) = 0,$$

we obtain

$$\begin{pmatrix} b(\hat{\lambda}_n) \\ \hat{\lambda}_n - \lambda_0 \end{pmatrix} = M_n^{-1}(\lambda^*, b^*) \begin{pmatrix} -Q_{1n}(\lambda_0, 0) \\ 0 \end{pmatrix} = M^{-1} \begin{pmatrix} -Q_{1n}(\lambda_0, 0) \\ 0 \end{pmatrix} + o_p(1),$$

where each element of λ^* lies between the corresponding ones of $\hat{\lambda}_n$ and λ_0, and each element of b^* lies between the corresponding ones of $b(\hat{\lambda}_n)$ and 0. Therefore,

$$\begin{aligned}
& l_n(r; \theta_0, \hat{\lambda}_n(\theta_0)) \\
= \; & -nQ'_{1n}(\lambda_0, 0) M_{11}^{-1}\{I + M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1}\}Q_{1n}(\lambda_0, 0) + o_p(1) \\
= \; & \{(-M_{11})^{-\frac{1}{2}}\sqrt{n}Q_{1n}(\lambda_0, 0)\}'\{I - (-M_{11})^{-\frac{1}{2}}M_{12}M_{22.1}^{-1}M_{21}(-M_{11})^{-\frac{1}{2}}\} \\
& \times\{(-M_{11})^{-\frac{1}{2}}\sqrt{n}Q_{1n}(\lambda_0, 0)\} + o_p(1).
\end{aligned}$$

It is easy to check that $(-M_{11})^{-\frac{1}{2}}\sqrt{n}Q_{1n}(\lambda_0, 0)$ converges in distribution to a standard multivariate normal distribution, and $I - (-M_{11})^{-\frac{1}{2}}M_{12}M_{22.1}^{-1}M_{21}(-M_{11})^{-\frac{1}{2}}$ is symmetric and idempotent with rank 1. Hence,

$$l_n(r; \theta_0, \hat{\lambda}_n(\theta_0)) \xrightarrow{d} \chi_1^2.$$

$\square$

In order to prove Proposition 3.2 and Theorem 3.2, we need the following lemma which is similar to Lemma 3.2.

Lemma 3.4. *Put*

$$\begin{aligned}
A_n(\lambda) &= \frac{1}{n}\sum_{t=1}^{n} \frac{\partial G\left(\frac{\theta_0/\sqrt{h_{n+1}(\lambda)} - \epsilon_t(\lambda)}{h}\right)}{\partial \lambda} \\
&= -\frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{\theta_0/\sqrt{h_{n+1}(\lambda)} - \epsilon_t(\lambda)}{h}\right)\left\{\frac{\partial \epsilon_t(\lambda)}{\partial \lambda} + \frac{\theta_0}{2h_{n+1}^{\frac{3}{2}}(\lambda)}\frac{\partial h_{n+1}(\lambda)}{\partial \lambda}\right\}.
\end{aligned}$$

Under conditions of Proposition 3.2, we have

$$\sup_{\lambda \in V_n} \|A_n(\lambda) - A_n(\lambda_0)\| = o_p(1). \tag{3.21}$$

Proof. Write

$$\begin{aligned} A_n(\lambda) &= -\frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{\theta_0/\sqrt{h_{n+1}(\lambda)} - \epsilon_t(\lambda)}{h}\right)\frac{\partial \epsilon_t(\lambda)}{\partial \lambda} \\ &\quad - \left\{\frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{\theta_0/\sqrt{h_{n+1}(\lambda)} - \epsilon_t(\lambda)}{h}\right)\right\}\frac{\theta_0}{2h_{n+1}^{\frac{3}{2}}(\lambda)}\frac{\partial h_{n+1}(\lambda)}{\partial \lambda} \\ &= -B_1(\lambda) - B_2(\lambda). \end{aligned}$$

Since $\sup_{\lambda \in V_n} |\frac{\theta_0}{\sqrt{h_{n+1}(\lambda)}} - \frac{\theta_0}{\sqrt{h_{n+1}(\lambda_0)}}| \to 0$ as $n \to \infty$, in a way similar to the proof of Lemma 2, we can show that

$$\sup_{\lambda \in V_n} \|B_1(\lambda) - B_1(\lambda_0)\| = o_p(1).$$

Also, similar to the proof of the uniform convergence of D_2 in Lemma 2, we have

$$\sup_{\lambda \in V_n} \|B_2(\lambda) - B_2(\lambda_0)\| = o_p(1).$$

Hence, Lemma 3.4 is proved. □

Proofs of Proposition 3.2 and Theorem 3.2. They are similar to the proofs of Proposition 3.1 and Theorem 3.1 by using Lemma 3.4 replacing Lemma 3.2. □

Table 3.1: Coverage probabilities for the confidence intervals of the conditional VaR based on the proposed empirical likelihood method for levels $\gamma = 0.90, 0.95$.

$(n, \alpha_1, z_1, \gamma)$	$\epsilon_t \sim N(0,1)$			$\epsilon_t \sim \sqrt{\frac{3}{5}}t(5)$		
	r=0.90	r=0.95	r=0.99	r=0.90	r=0.95	r=0.99
$(1000, 0.7, 0, 0.90)$	0.9070	0.9120	0.9242	0.8816	0.8880	0.9252
$(3000, 0.7, 0, 0.90)$	0.9088	0.8980	0.9074	0.9030	0.8932	0.9060
$(1000, 0.9, 0, 0.90)$	0.9040	0.9106	0.9278	0.8694	0.8798	0.9396
$(3000, 0.9, 0, 0.90)$	0.9010	0.9042	0.9056	0.8870	0.8862	0.9110
$(1000, 0.7, 1, 0.90)$	0.900	0.9018	0.9392	0.8928	0.8904	0.9408
$(3000, 0.7, 1, 0.90)$	0.9048	0.9052	0.9064	0.8980	0.9044	0.9066
$(1000, 0.9, 1, 0.90)$	0.9030	0.9088	0.9454	0.8772	0.8966	0.9468
$(3000, 0.9, 1, 0.90)$	0.9038	0.8988	0.9122	0.8918	0.8952	0.9276
$(1000, 0.7, 0, 0.95)$	0.9526	0.9584	0.9774	0.9358	0.9406	0.9766
$(3000, 0.7, 0, 0.95)$	0.9536	0.9504	0.9532	0.9514	0.9468	0.9542
$(1000, 0.9, 0, 0.95)$	0.9546	0.9576	0.9836	0.9364	0.9356	0.9788
$(3000, 0.9, 0, 0.95)$	0.9488	0.9514	0.9534	0.9388	0.9386	0.9598
$(1000, 0.7, 1, 0.95)$	0.949	0.9564	0.9862	0.9420	0.9420	0.9876
$(3000, 0.7, 1, 0.95)$	0.9504	0.9496	0.9550	0.9494	0.9524	0.9554
$(1000, 0.9, 1, 0.95)$	0.9498	0.9542	0.9872	0.9380	0.9470	0.9856
$(3000, 0.9, 1, 0.95)$	0.9520	0.9474	0.9584	0.9458	0.9460	0.9656

CHAPTER IV

CONDITIONAL VALUE-AT-RISK IN HETEROSCEDASTIC REGRESSION MODELS

In this chapter, we continue to discuss empirical likelihood for estimating conditional Value-at-Risk but with different underlying model–nonparametric regression model. Nonparametric regression models have been studied well including estimating the conditional mean function, the conditional variance function and the distribution function of errors. In addition, empirical likelihood methods have been proposed to construct confidence intervals for the conditional mean and variance. Motivated by applications in risk management, we propose an empirical likelihood method for constructing a confidence interval for the pth conditional Value-at-Risk based on the nonparametric regression model. A simulation study shows the advantages of the proposed method. The content of this chapter is based on Z. Li, Y. Gong, and L. Peng (2011), Empirical likelihood intervals for conditional Value-at-Risk in heteroscedastic regression models. *Scandinavian Journal of Statistics,* Accepted.

4.1 Introduction

Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent bivariate observations sampled from the nonparametric regression model

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i, \tag{4.1}$$

where $m(\cdot) = E(Y_1|X_1 = \cdot)$ and $\sigma^2(\cdot) = Var(Y_1|X_1 = \cdot)$ are the conditional mean function and the conditional variance function, respectively. The random errors $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed random variables with zero mean and unit variance, and the sequence $\{\epsilon_i\}$ is independent of the sequence $\{X_i\}$.

Estimating the conditional mean function $m(x)$ has been studied extensively in the literature; see Fan and Gijbels (1996). References on estimating the conditional variance $\sigma^2(x)$ include Fan and Yao (1998), Yu and Jones (2004), Cai and Wang (2008), Wang, Brown, Cai and Levine (2008), Cai, Levine and Wang (2009) and Chen, Cheng and Peng (2009). Recently, study on estimating the error distribution and goodness-of-fit test for errors has received some attention; see Akritas and Van Keilegom (2001), Einmahl and Van Keilegom (2008) and Neumeyer and Van Keilegom (2010).

Although nonparametric conditional quantile estimation without assuming (4.1) has been studied well in the literature (see Bhattacharya and Kong (2007), Dette, Neumeyer and Pilz (2005), Dette and Volgushev (2008), Müller and Schmitt (1988), and Park and Park (2006)), we consider the interval estimation for the conditional Value-at-Risk based on the model (4.1) in this chapter.

The pth conditional VaR $\theta_p(x)$ of the conditional distribution of Y_1 given $X_1 = x$ is defined as

$$\theta_p(x) = \inf\{y : P(Y_1 \leq y | X_1 = x) \geq p\}, \tag{4.2}$$

where $0 < p < 1$. Using model (4.1), we have

$$\theta_p(x) = m(x) + \sigma(x) Q_\epsilon(p), \tag{4.3}$$

where $Q_\epsilon(p)$ is the pth quantile of the distribution of ϵ_1. Hence, a simple estimator for $\theta_p(x)$ is to replace $m(x), \sigma(x)$ and $Q_\epsilon(p)$ in (4.3) by corresponding estimators. Without doubt, this conditional quantile estimate based on (4.1) should perform better than the general nonparametric conditional quantile estimates if model (4.1) is correct. Since the estimator for $Q_\epsilon(p)$ has a rate of convergence $n^{-1/2}$ and estimators for $m(x)$ and $\sigma(x)$ have a rate of convergence $(nh)^{-1/2}$ with $h = h(n) \to 0$, the limit of the proposed estimator for $\theta_p(x)$ is generally determined by the joint limit of estimators for $m(x)$ and $\sigma(x)$. For constructing a confidence interval for the conditional VaR

$\theta_p(x)$, one has to estimate the asymptotic variance by deriving the joint limit of the estimators for $m(x)$ and $\sigma(x)$. In order to avoid estimating the asymptotic variance explicitly, we investigate the possibility of employing empirical likelihood method.

Applying the empirical likelihood method to the model (4.1) is not new at all. For example, Chen and Qin (2000) constructed an empirical likelihood confidence interval for the conditional mean function, Chan, Peng and Zhang (2010) constructed confidence intervals for the conditional variance function, Chen and Van Keilegom (2009a) studied the empirical likelihood test for multiresponse regression, Chen and Van Keilegom (2009b) gave an excellent review on empirical likelihood methods based on regression models. However, how to construct an empirical likelihood confidence interval for the conditional VaR based on the model (4.1) is not available in the literature.

4.2 Main results

First we estimate the conditional mean function $m(x)$ and the conditional second moment $M(x)$ by the local linear technique (see Fan and Gijbels (1996)). That is, by solving

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \sum_{i=1}^{n} \{Y_i - a - b(X_i - x)\}^2 K(\frac{X_i - x}{h}) \tag{4.4}$$

and

$$(\hat{c}, \hat{d}) = \arg\min_{c,d} \sum_{i=1}^{n} \{Y_i^2 - c - d(X_i - x)\}^2 K(\frac{X_i - x}{h}), \tag{4.5}$$

where K is a density function with support $[-1, 1]$ and $h = h(n) > 0$ is a bandwidth, the estimators for $m(x)$ and $M(x)$ are defined as $\hat{m}(x) = \hat{a}$ and $\hat{M}(x) = \hat{c}$, respectively. Further we define the estimator for $\sigma^2(x)$ as $\hat{\sigma}^2(x) = \hat{M}(x) - (\hat{m}(x))^2$.

After estimating $m(x)$ and $\sigma(x)$, we can estimate the errors by

$$\hat{\epsilon}_i = \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)} \quad \text{for } i = 1, \ldots, n. \tag{4.6}$$

So the estimator $\hat{Q}_\epsilon(p)$ for $Q_\epsilon(p)$ can be chosen as the pth sample quantile of $\hat{\epsilon}_1, \cdots, \hat{\epsilon}_n$.

Note that the local linear estimators for $m(x)$ and $M(x)$ are the same as the solution to the following estimating equations,

$$\frac{1}{n}\sum_{i=1}^{n} W_i(Y_i - m(x)) = 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} W_i(Y_i^2 - M(x)) = 0,$$

where $W_i = K(\frac{X_i - x}{h})\left\{S_{n,2} - \frac{(X_i - x)S_{n,1}}{h}\right\}$ and $S_{n,l} = (nh)^{-1}\sum_{i=1}^{n} K(\frac{X_i - x}{h})\frac{(X_i - x)^l}{h^l}$ for $l = 1, 2$. By writing $M(x) = \{\frac{\theta_p(x) - m(x)}{Q_\epsilon(p)}\}^2 + \{m(x)\}^2$, these formulations allow us to employ the method in Qin and Lawless (1994) to construct an empirical likelihood confidence interval for $\theta_p(x)$ as follows.

Let $(p_1, \ldots, p_n)$ be a probability vector, i.e., $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$ for $1 \leq i \leq n$. We use θ and α to denote $\theta_p(x)$ and $m(x)$, respectively. Then, the empirical likelihood ratio is $R(\theta, \alpha) = \sup\{\prod_{i=1}^{n}(np_i)\}$ subject to the constraints

$$\begin{cases} \sum_{i=1}^{n} p_i W_i (Y_i - \alpha) = 0 \\ \sum_{i=1}^{n} p_i W_i \left\{ Y_i^2 - \left(\frac{\theta - \alpha}{\hat{Q}_\epsilon(p)}\right)^2 - \alpha^2 \right\} = 0. \end{cases}$$

Note that we use $\hat{Q}_\epsilon(p)$ in the above constraints instead of $Q_\epsilon(p)$. The reason is that the rate of convergence for estimating $Q_\epsilon(p)$ is faster than that for estimating $m(x)$.

Write

$$V_{1i}(\alpha) = W_i\{Y_i - \alpha\} \quad \text{and} \quad V_{2i}(\alpha) = V_{2i}(\alpha, \theta) = W_i \left\{ Y_i^2 - \left(\frac{\theta - \alpha}{\hat{Q}_\epsilon(p)}\right)^2 - \alpha^2 \right\}.$$

By the argument of Lagrange multipliers, we find that $\log R(\theta, \alpha)$ is maximized at

$$p_i = \frac{1}{n}\frac{1}{1 + b_1 V_{1i}(\alpha) + b_2 V_{2i}(\alpha)}, \quad i = 1, \ldots, n,$$

which gives the log empirical likelihood ratio

$$l_n(\theta, \alpha) = -2\log R(\theta, \alpha) = 2\sum_{i=1}^{n} \log\{1 + b_1 V_{1i}(\alpha) + b_2 V_{2i}(\alpha)\}, \tag{4.7}$$

where $b_1 = b_1(\alpha)$ and $b_2 = b_2(\alpha)$ satisfy the following equations:

$$Q_{1n}(\alpha, b_1, b_2) := \frac{1}{nh}\sum_{i=1}^{n} \frac{V_{1i}(\alpha)}{1 + b_1 V_{1i}(\alpha) + b_2 V_{2i}(\alpha)} = 0, \tag{4.8}$$

$$Q_{2n}(\alpha, b_1, b_2) := \frac{1}{nh}\sum_{i=1}^{n}\frac{V_{2i}(\alpha)}{1 + b_1V_{1i}(\alpha) + b_2V_{2i}(\alpha)} = 0. \tag{4.9}$$

Since our interest is θ, we minimize $l_n(\theta, \alpha)$ with respect to α and get the following profile log empirical likelihood ratio

$$l_n(\theta) = l_n(\theta, \hat{\alpha}(\theta)) = 2\sum_{i=1}^{n}\log\{1 + b_1V_{1i}(\hat{\alpha}) + b_2V_{2i}(\hat{\alpha})\}, \tag{4.10}$$

where $\hat{\alpha}(\theta) = \arg\min_\alpha l_n(\theta, \alpha)$.

Before deriving the asymptotic limit of $l_n(\theta)$, we list some regularity conditions.

Assumption 4.1. *(i) K is a symmetric continuously differentiable density function with support on $[-1, 1]$ and $K(\pm 1) = 0$. Denote $\tau_1 = \int_{-1}^{1} K^2(u)du$ and $\tau_2 = \int_{-1}^{1} u^2K(u)du$;*

(ii) $h = h(n) \to 0$, $n^{s-4}h^s(nh)^{-\iota} \to \infty$ for some $s - 6 < \iota < s - \frac{16}{3}$, and $nh^4 \to 0$, as $n \to \infty$, where $s \geq 6$, and $E|Y_1|^s < \infty$;

(iii) Denote F_X and f as the distribution and density of X_1 which are defined on some compact set R_X in R^1. All the derivatives of F_X up to order 3 exist on the interior of R_X and they are uniformly continuous, and $\inf_{x\in R_X} f(x) > 0$;

(iv) All the derivatives of m and σ up to order 3 exist on the interior of R_X and they are uniformly continuous, and $\inf_{x\in R_X} \sigma(x) > 0$;

(v) Define F_ϵ and f_ϵ as the distribution and density of ϵ_1. Assume F_ϵ is twice continuously differentiable and $\sup_y |y^2 f_\epsilon'(y)| < \infty$.

Remark 4.1. *i) By taking $s = 6$, Assumption 1(ii) can be replaced by*

(ii)′ As $n \to \infty, h = h(n) \to 0, nh^{3+\eta} \to \infty$ for some $\eta > 0$ and $nh^4 \to 0$.

ii) Assumption 4.1(i), (iii)-(v) and $E|Y_1|^6 < \infty$ in (ii) come from the regularity conditions in Neumeyer and Van Keilegom (2010), which are employed to ensure the $n^{-1/2}$ rate of convergence for estimating the distribution function of ϵ_1.

iii) We can employ different bandwidths for $\hat{m}$ and $\hat{\sigma}$ in $\hat{Q}_\epsilon(p)$ and $l_n(\theta)$.

Denote θ_0 and $\alpha_0(= m(x))$ as the true values of θ and α, i.e, the true values of $\theta_p(x)$ and $m(x)$. First we have the following proposition on the existence of $\hat{\alpha}(\theta_0)$.

Proposition 4.1. *(Li, Gong and Peng, 2011) Under Assumption 4.1, with probability tending to one, we have* $|\hat{\alpha}(\theta_0) - \alpha_0| < (nh)^{-1/2+\delta}$, *for some* $0 < \delta < \frac{\iota}{2s}$. *Moreover,* $\hat{\alpha}(\theta_0)$, b_1 *and* b_2 *in* (4.10) *satisfy the equations* (4.8), (4.9) *and*

$$Q_{3n}(\alpha, b_1, b_2) := \frac{1}{nh} \sum_{i=1}^{n} \frac{b_1 V_{1i}'(\alpha) + b_2 V_{2i}'(\alpha)}{1 + b_1 V_{1i}(\alpha) + b_2 V_{2i}(\alpha)} = 0. \tag{4.11}$$

With the above estimator $\hat{\alpha}$, the following Wilks' theorem holds for the proposed profile log empirical likelihood ratio.

Theorem 4.1. *(Li, Gong and Peng, 2011) Under Assumption 4.1, we have* $l_n(\theta_0, \hat{\alpha}(\theta_0)) \xrightarrow{d} \chi_1^2$ *as* $n \to \infty$.

Based on the above theorem, we can construct an asymptotic confidence interval with level γ for the *pth* conditional VaR $\theta_p(x)$ as

$$I_\gamma = \{\bar{\theta} : l_n(\bar{\theta}, \hat{\alpha}(\bar{\theta})) \le \chi_{1,\gamma}^2\},$$

where $\chi_{1,\gamma}^2$ is the γth quantile of χ_1^2.

Remark 4.2. *Although the above theorem holds when X is a d-dimensional random vector for some $d > 1$, the proposed procedure suffers from the curse of dimensionality as model (4.1) does.*

4.3 Simulation

In this section, a simulation study is carried out to investigate the accuracy of the proposed empirical likelihood confidence intervals for the conditional VaR and compare with the normal approximation method. We consider the following two models:

(i) $m(x) = x$, $\sigma(x) = \exp(-x) + 0.5$;

(ii) $m(x) = x + 2\exp(-16x)$, $\sigma(x) = 2\exp(-x) + 0.2$.

The design variable X_1 and the random error ϵ_1 have a uniform distribution on $[0,1]$ and the standard normal distribution, respectively. For each model, we consider the pth conditional VaR at $X_1 = 0.25$ and 0.5 for $p = 0.95$ with confidence level $\gamma = 0.9$ and 0.95. In each of the settings, we draw $1,000$ random samples with sample size $n = 200$ and $n = 1000$ from the above two models. We employ the kernel function $K(x) = \frac{15}{16}(1-x^2)^2 I(|x| \le 1)$ and the bandwidth $h = 0.5n^{-7/24}$ and $h = n^{-7/24}$.

For computing the coverage probability of the normal approximation method, we employ the jackknife variance estimation for the pth conditional VaR estimator $\hat{\theta}_p(x) = \hat{m}(x) + \hat{\sigma}(x)\hat{Q}_\epsilon(p)$. That is, we denote $\hat{m}_{-i}(x)$ and $\hat{\sigma}_{-i}(x)$ as the corresponding estimators based on the sample $\{(X_1,Y_1),\cdots,(X_{i-1},Y_{i-1}),(X_{i+1},Y_{i+1}),\cdots,(X_n,Y_n)\}$. Then we have $\hat{\theta}_{p,-i}(x) = \hat{m}_{-i}(x) + \hat{\sigma}_{-i}(x)\hat{Q}_\epsilon(p)$ for $i = 1,\cdots,n$. Note that we keep $\hat{Q}_\epsilon(p)$ in the Jackknife sample since it does not contribute to the limit of $\hat{\theta}_p(x)$. Based on the Jackknife sample $\{\hat{\theta}_{p,-i}(x)\}_{i=1}^n$, we estimate the asymptotic variance of $\hat{\theta}_p(x)$ by

$$\hat{\upsilon}_J = \frac{n-1}{n}\sum_{i=1}^n \{\hat{\theta}_{p,-i}(x) - \frac{1}{n}\sum_{j=1}^n \hat{\theta}_{p,-j}(x)\}^2.$$

Therefore a normal approximation based confidence interval for $\theta_p(x)$ with confidence level γ can be obtained as

$$I_\gamma^N = [\hat{\theta}_p(x) - \sqrt{\hat{\upsilon}_J} z_{\gamma/2}, \quad \hat{\theta}_p(x) + \sqrt{\hat{\upsilon}_J} z_{\gamma/2}],$$

where $z_{\gamma/2}$ satisfies $P(N(0,1) > z_{\gamma/2}) = \gamma/2$.

We also consider a bootstrap calibration for the proposed empirical likelihood method. More specifically, we draw 1000 bootstrap samples with size n from the standardized estimated errors $\bar{\epsilon}_i = \{\hat{\epsilon}_i - \frac{1}{n}\sum_{j=1}^n \hat{\epsilon}_j\}/\{\frac{1}{n}\sum_{j=1}^n(\hat{\epsilon}_j - \frac{1}{n}\sum_{l=1}^n \hat{\epsilon}_l)^2\}^{1/2}$ for $i = 1,\cdots,n$, where $\hat{\epsilon}_i's$ are defined in (4.6). Denote them by $\hat{\epsilon}_i^{j*}, i = 1,\cdots,n, j = 1,\cdots,300$. Using these bootstrapped errors, we generate

$$Y_i^{j*} = \hat{m}(X_i) + \hat{\sigma}(X_i)\hat{\epsilon}_i^{j*},$$

where $\hat{m}(x)$ and $\hat{\sigma}(x)$ are defined right before (4.6). Then, for each $j = 1, \cdots, 300$, we recalculate the empirical likelihood ratio $l_n(\hat{\theta}_p(x), \hat{\alpha}(\hat{\theta}_p(x)))$ based on the bootstrap sample $(X_i, Y_i^{j*}), i = 1, \cdots, n$. That is, we obtain 1000 such empirical likelihood ratios and denote the $[1000\gamma] - th$ largest value by d^*. Hence the calibrated empirical likelihood confidence interval is defined as

$$I_\gamma^C = \{\theta : l_n(\theta, \hat{\alpha}(\theta) \leq d^*\}.$$

In Tables 4.1-4.2, we report coverage probabilities for the confidence intervals of the $0.95th$ conditional VaR with levels $\gamma = 0.9$ and 0.95 based on the proposed empirical likelihood method, the bootstrap calibration method and the normal approximation method. From these two tables, we observe that i) the proposed empirical likelihood method performs slightly worse than the normal approximation method; ii) the calibration method produces most accurate confidence intervals; iii) coverage accuracy improves for three methods when the sample size increases. As usual, how to choose bandwidth in terms of coverage accuracy rather than mean squared errors remains difficult both theoretically and practically.

4.4 Proofs

For simplicity, we first introduce some notations. For $j = 1, 2$, define

$$\begin{cases} \bar{V}_j^1(\alpha) = \frac{1}{nh}\sum_{i=1}^n V_{ji}(\alpha) \\ \bar{V}_j^2(\alpha) = \frac{1}{nh}\sum_{i=1}^n V_{ji}^2(\alpha) \\ \bar{V}_3(\alpha) = \frac{1}{nh}\sum_{i=1}^n V_{1i}(\alpha)V_{2i}(\alpha) \\ \bar{W} = \frac{1}{nh}\sum_{i=1}^n W_i \\ \bar{W}^* = \frac{1}{nh}\sum_{i=1}^n W_i^2(Y_i - \alpha_0), \end{cases}$$

and $\bar{w}_{ijk} = (nh)^{-1}\sum_{l=1}^{n}(\frac{X_l - x}{h})^i K^j(\frac{X_l - x}{h})(Y_l - \alpha_0)^k$. Hence

$$\begin{cases} \bar{V}_1^1(\alpha_0) & = \bar{w}_{210}\bar{w}_{011} - \bar{w}_{110}\bar{w}_{111} \\ \bar{V}_2^1(\alpha_0) & = \bar{w}_{210}\bar{w}_{012} - \bar{w}_{110}\bar{w}_{112} + 2\alpha_0\bar{w}_{210}\bar{w}_{011} - 2\alpha_0\bar{w}_{110}\bar{w}_{111} \\ & \quad - \left(\frac{\theta_0 - \alpha_0}{\hat{Q}_\epsilon(p)}\right)^2 \bar{w}_{210}\bar{w}_{010} + \left(\frac{\theta_0 - \alpha_0}{\hat{Q}_\epsilon(p)}\right)^2 \bar{w}_{110}^2 \\ \bar{W} & = \bar{w}_{210}\bar{w}_{010} - \bar{w}_{110}^2 \\ \bar{W}^* & = \bar{w}_{210}^2\bar{w}_{021} + \bar{w}_{110}^2\bar{w}_{221} - 2\bar{w}_{110}\bar{w}_{210}\bar{w}_{121}. \end{cases}$$

By the central limit theorem and Delta method, we can show that

$$\begin{cases} \bar{V}_r^1(\alpha_0) - E\{\bar{V}_r^1(\alpha_0)\} = O_p((nh)^{-1/2}), \qquad r = 1, 2 \\ \bar{W} - E\{\bar{W}\} = O_p((nh)^{-1/2}) \\ \bar{W}^* - E\{\bar{W}^*\} = O_p((nh)^{-1/2}). \end{cases} \tag{4.12}$$

It is easy to check that

$$E\{\bar{W}\} = f^2(x)\tau_2 + O(h^2), \qquad E\{\bar{W}^*\} = O(h^2). \tag{4.13}$$

Before proving Proposition 4.1 and Theorem 4.1, we need a few lemmas.

Lemma 4.1. *Under Assumption 4.1, we have*

$$\begin{cases} E\bar{V}_1^1(\alpha) & = f^2(x)\{m(x) - \alpha\}\tau_2 + O(h^2) \\ E\{\bar{V}_2^1(\alpha)\} & = f^2(x)\{m^2(x) + \sigma^2(x) - T\}\tau_2 + O(n^{-1/2} + (nh)^{-1}) \\ E\{\bar{V}_1^2(\alpha)\} & = f^3(x)\{(m(x) - \alpha)^2 + \sigma^2(x)\}\tau_1\tau_2^2 + O(h + (nh)^{-1}) \\ E\{\bar{V}_2^2(\alpha)\} & = f^3(x)\{m^4(x) + \sigma^4(x)E\epsilon^4 + 6m^2(x)\sigma^2(x) + 4m(x)\sigma^3(x)E\epsilon^3 \\ & \quad -2T[m^2(x) + \sigma^2(x)] + T^2\}\tau_1\tau_2^2 + O(n^{-1/2} + h + (nh)^{-1}) \\ E\{\bar{V}_3(\alpha)\} & = f^3(x)\{m^3(x) + \sigma^3(x)E\epsilon^3 + 3m(x)\sigma^2(x) - Tm(x) - \alpha[m^2(x) + \sigma^2(x) - T]\} \\ & \quad \times\tau_1\tau_2^2 + O(n^{-1/2} + h + (nh)^{-1}), \end{cases}$$

where $T = \left(\frac{\theta_0 - \alpha}{Q_\epsilon(p)}\right)^2 + \alpha^2$ *and* $O(\cdot)$ *holds uniformly for* α *in a compact set including* $\alpha_0 = m(x)$.

Proof. We only show $E\{\bar{V}_1^1(\alpha)\}$ since the rest can be done in a similar way. By definition, we have

$$
\begin{aligned}
& E\{\bar{V}_1^1(\alpha)\} \\
= \; & E\{S_{n,2}(nh)^{-1}\sum_{i=1}^{n} K(\frac{X_i - x}{h})(Y_i - \alpha) - S_{n,1}(nh)^{-1}\sum_{i=1}^{n} K(\frac{X_i - x}{h})(\frac{X_i - x}{h})(Y_i - \alpha)\} \\
= \; & (nh)^{-2}E\{\sum_{i\neq j}(\frac{X_i - x}{h})^2 K(\frac{X_i - x}{h})K(\frac{X_j - x}{h})(Y_j - \alpha) \\
& - \sum_{i\neq j}(\frac{X_i - x}{h})K(\frac{X_i - x}{h})(\frac{X_j - x}{h})K(\frac{X_j - x}{h})(Y_j - \alpha)\}. \qquad (4.14)
\end{aligned}
$$

Clearly,

$$
\begin{aligned}
& E\{(\frac{X_i - x}{h})K(\frac{X_i - x}{h})(Y_i - \alpha)\} \\
= \; & E\{(\frac{X_i - x}{h})K(\frac{X_i - x}{h})[m(X_i) - \alpha]\} \\
= \; & \int_{-\infty}^{\infty}(\frac{y - x}{h})K(\frac{y - x}{h})[m(y) - \alpha]f(y)dy \\
= \; & h\int_{-1}^{1} uK(u)[m(x + hu) - \alpha]f(x + hu)du \\
= \; & h\int_{-1}^{1} uK(u)[m(x) - \alpha + hum'(x^*)][f(x) + huf'(x^{**})]du \\
= \; & O(h^2),
\end{aligned}
$$

where x^* and x^{**} lie in $[x - h, x + h]$.

Similarly, we have

$$E\{(\frac{X_i - x}{h})^2 K(\frac{X_i - x}{h})\} = hf(x)\tau_2 + O(h^3),$$

$$E\{K(\frac{X_j - x}{h})(Y_j - \alpha)\} = hf(x)[m(x) - \alpha] + O(h^3)$$

and

$$E\{(\frac{X_i - x}{h})K(\frac{X_i - x}{h})\} = O(h^2).$$

Since (X_i, Y_i) and (X_j, Y_j) are independent for $i \neq j$, it follows from (4.14) that

$$E\{\bar{V}_1^1(\alpha)\} = f^2(x)[m(x) - \alpha]\tau_2 + O(h^2).$$

$\square$

Lemma 4.2. *Under Assumption 4.1, we have*

$$\begin{cases} \bar{V}_1^1(\alpha) = O_p((nh)^{-1/2+\delta}) \\ \bar{V}_2^1(\alpha) = O_p((nh)^{-1/2+\delta}) \\ \bar{V}_1^2(\alpha) = E\{\bar{V}_1^2(\alpha_0)\} + O_p(h + (nh)^{-1+2\delta}) \\ \bar{V}_2^2(\alpha) = E\{\bar{V}_2^2(\alpha_0)\} + O_p(h + (nh)^{-1+2\delta}) \\ \bar{V}_3(\alpha) = E\{\bar{V}_3(\alpha_0)\} + O_p(h + (nh)^{-1+2\delta}), \end{cases}$$

where $O_p(\cdot)$ holds uniformly for $\alpha \in \{\alpha : |\alpha - \alpha_0| \le (nh)^{-1/2+\delta}\}$.

Proof. By definition, we have

$$\begin{aligned} & \bar{V}_2^1(\alpha) - \bar{V}_2^1(\alpha_0) \\ &= \tfrac{1}{nh} \textstyle\sum_{i=1}^n W_i \left\{ \left(\frac{\theta_0 - \alpha_0}{\hat{Q}_\epsilon(p)} \right)^2 - \left(\frac{\theta_0 - \alpha}{\hat{Q}_\epsilon(p)} \right)^2 + \alpha_0^2 - \alpha^2 \right\} \\ &= \tfrac{1}{nh} \textstyle\sum_{i=1}^n W_i \left\{ \frac{2\theta_0 - \alpha_0 - \alpha}{\hat{Q}_\epsilon(p)} \frac{\alpha - \alpha_0}{\hat{Q}_\epsilon(p)} + (\alpha_0 - \alpha)(\alpha_0 + \alpha) \right\} \\ &= \bar{W} O_p(|\alpha - \alpha_0|) \\ &= O_p((nh)^{-1/2+\delta}). \end{aligned}$$

It follows from (4.12) and Lemma 4.1 that

$$\bar{V}_2^1(\alpha_0) = O_p((nh)^{-1/2} + n^{-1/2} + (nh)^{-1}) = O_p((nh)^{-1/2}).$$

So

$$\bar{V}_2^1(\alpha) = O_p((nh)^{-1/2+\delta}).$$

The rest can be shown similarly. □

Lemma 4.3. *For $b_1 = b_1(\alpha)$ and $b_2 = b_2(\alpha)$ satisfying (4.8) and (4.9), we have, uniformly for $\alpha \in \{\alpha : |\alpha - \alpha_0| \le (nh)^{-1/2+\delta}\}$,*

$$b_i = O_p((nh)^{-1/2+\delta}), \quad i = 1, 2.$$

Proof. Denote $V_i(\alpha) = (V_{1i}(\alpha), V_{2i}(\alpha))^T$, $b = (b_1, b_2)^T$ and $b = \rho u$ with $\|u\| = 1$.

From (4.8) and (4.9), we get

$$
\begin{aligned}
0 &= \|\frac{1}{nh}\sum_{i=1}^{n}\frac{V_i(\alpha)}{1+b^TV_i(\alpha)}\| \\
&\geq |u^T\frac{1}{nh}\sum_{i=1}^{n}\frac{V_i(\alpha)}{1+b^TV_i(\alpha)}| \\
&\geq \frac{\rho u^TS_n(\alpha)u}{1+\rho Z_n(\alpha)} - |\frac{1}{nh}\sum_{i=1}^{n}u^TV_i(\alpha)|, \qquad (4.15)
\end{aligned}
$$

where

$$
S_n(\alpha) = (nh)^{-1}\sum_{i=1}^{n}V_i(\alpha)V_i^T(\alpha) = \begin{pmatrix} \bar{V}_1^2(\alpha) & \bar{V}_3(\alpha) \\ \bar{V}_3(\alpha) & \bar{V}_2^2(\alpha) \end{pmatrix} \quad \text{and} \quad Z_n(\alpha) = \max_{1\leq i\leq n}\|V_i(\alpha)\|.
$$

By Lemma 4.2, we have

$$
S_n(\alpha) = \begin{pmatrix} \bar{V}_1^2(\alpha_0) & \bar{V}_3(\alpha_0) \\ \bar{V}_3(\alpha_0) & \bar{V}_2^2(\alpha_0) \end{pmatrix} + o_p(1) \xrightarrow{p} S_0, \qquad (4.16)
$$

and S_0 is positive definite.

Since $S_{n,l} - E(S_{n,l}) = O_p((nh)^{-1/2})$ for $l = 1, 2$, $E(S_{n,1}) = O(h)$ and $E(S_{n,2}) = O(1)$, it follows from Assumption 4.1(ii) that

$$
\begin{aligned}
Z_n(\alpha) = \max_{1\leq i\leq n}\|V_i(\alpha)\| &\leq \max_{1\leq i\leq n}|V_{1i}(\alpha)| + \max_{1\leq i\leq n}|V_{2i}(\alpha)| \\
&\leq (\max_{1\leq i\leq n}|Y_i| + \max_{1\leq i\leq n}|Y_i^2|)O_p(1) + O_p(1) \\
&= O_p(n^{2/s}). \qquad (4.17)
\end{aligned}
$$

Moreover, we have $V(\alpha) := (nh)^{-1}\sum_{i=1}^{n}V_i(\alpha) = O_p((nh)^{-1/2+\delta})$.

Hence, it follows from (4.15) and Assumption 1(ii) that $\|b\| = O_p((nh)^{-1/2+\delta})$ uniformly for $\alpha \in \{\alpha : |\alpha - \alpha_0| \leq (nh)^{-1/2+\delta}\}$. □

Proof of Proposition 4.1. Denote $\gamma_i = b^TV_i(\alpha)$. It follows from (4.17) and Assumption 1 (ii) that

$$
\max_{1\leq i\leq n}|\gamma_i| \leq \|b^T\|Z_n(\alpha) = o_p(1).
$$

Applying Taylor expansions to (4.8) and (4.9), we have

$$\begin{aligned}
0 &= V(\alpha) - S_n(\alpha)b + (nh)^{-1}\sum_{i=1}^{n}\frac{\gamma_i^2 V_i(\alpha)}{1+\gamma_i} \\
&\leq V(\alpha) - S_n(\alpha)b + \frac{\|b\|\max_{1\leq i\leq n}|\gamma_i|}{1-\max_{1\leq i\leq n}|\gamma_i|}(nh)^{-1}\sum_{i=1}^{n}|V_i(\alpha)|^2 \\
&\leq V(\alpha) - S_n(\alpha)b + o_p((nh)^{-1/2+\delta}).
\end{aligned}$$

Thus,

$$b = S_n^{-1}(\alpha)V(\alpha) + o_p((nh)^{-1/2+\delta}).$$

Put $\alpha_1 = \alpha_0 + (nh)^{-1/2+\delta}$. By expanding the empirical likelihood ratio function, we have

$$\begin{aligned}
l_n(\theta_0,\alpha_1) &= 2\sum_{i=1}^{n} b^T(\alpha_1)V_i(\alpha_1) - \sum_{i=1}^{n}\{b^T(\alpha_1)V_i(\alpha_1)\}^2 + o_p((nh)^{2\delta}) \\
&= nhV^T(\alpha_1)S_n^{-1}(\alpha_1)V(\alpha_1) + o_p((nh)^{2\delta}) \\
&= nh\{V(\alpha_0) + \frac{dV(\alpha^*)}{d\alpha}(nh)^{-1/2+\delta}\}^T S_n^{-1}(\alpha_1)\{V(\alpha_0) + \frac{dV(\alpha^*)}{d\alpha}(nh)^{-1/2+\delta}\} \\
&\quad + o_p((nh)^{2\delta}) \\
&= nh\{O_p((nh)^{-1/2}) + E\{\frac{dV(\alpha_0)}{d\alpha}\}(nh)^{-1/2+\delta}\}^T \\
&\quad \times S_0^{-1}\{O_p((nh)^{-1/2}) + E\{\frac{dV(\alpha_0)}{d\alpha}\}(nh)^{-1/2+\delta}\} + o_p((nh)^{2\delta}) \\
&\geq (c-\epsilon)(nh)^{2\delta},
\end{aligned}$$

where $c > 0$ is the smallest eigenvalue of

$$E\{\frac{dV(\alpha_0)}{d\alpha}\}^T S_0^{-1} E\{\frac{dV(\alpha_0)}{d\alpha}\},$$

with $\frac{dV(\alpha_0)}{d\alpha} = (\frac{d\bar{V}_1^1(\alpha_0)}{d\alpha}, \frac{dV_2^1(\alpha_0)}{d\alpha})^T$,

$$\begin{aligned}
E\{\frac{d\bar{V}_1^1(\alpha_0)}{d\alpha}\} &= E(-\bar{W}) = -f^2(x)\tau_2 + o_p(1), \\
E\{\frac{d\bar{V}_2^1(\alpha_0)}{d\alpha}\} &= E\{2\left\{\frac{\theta_0-\alpha_0}{\hat{Q}_\epsilon^2(p)} - \alpha_0\right\}\bar{W}\} = 2\left\{\frac{\sigma(x)}{Q_\epsilon(p)} - \alpha_0\right\}f^2(x)\tau_2 + o_p(1).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
l_n(\theta_0,\alpha_0) &= nhV(\alpha_0)^T S_n^{-1}(\alpha_0)V(\alpha_0) + o_p(1) \\
&= O_p(1).
\end{aligned}$$

Hence, $l_n(\theta_0, \alpha_0 + (nh)^{-1/2+\delta}) > l_n(\theta_0, \alpha_0)$ with probability tending to one as $n \to \infty$. Similarly, we can obtain $l_n(\theta_0, \alpha_0 - (nh)^{-1/2+\delta}) > l_n(\theta_0, \alpha_0)$ with probability tending to one as $n \to \infty$.

Since $l_n(\theta_0, \alpha)$ is a continuous function of α in $[\alpha_0 - (nh)^{-1/2+\delta}, \alpha_0 + (nh)^{-1/2+\delta}]$, $l_n(\theta_0, \alpha)$ attains its minimum at some interior point of $[\alpha_0 - (nh)^{-1/2+\delta}, \alpha_0 + (nh)^{-1/2+\delta}]$, say, $\hat{\alpha}$. By differentiating (4.7), we know that $\hat{\alpha}$, $b_1(\hat{\alpha})$ and $b_2(\hat{\alpha})$ satisfy equations (4.8)-(4.11). $\square$

Proof of Theorem 4.1. Put $\hat{\alpha} = \hat{\alpha}(\theta_0)$, $\hat{b}_i = b_i(\hat{\alpha})$, for $i = 1, 2$. Remember that, from (4.8), (4.9) and (4.11),

$$Q_{jn}(\hat{\alpha}, \hat{b}_1, \hat{b}_2) = 0, \quad \text{for } j = 1, 2, 3.$$

For simplicity, we denote $Q_{1,2n}(\alpha, b_1, b_2) = (Q_{1n}(\alpha, b_1, b_2), Q_{2n}(\alpha, b_1, b_2))^T$ and $b = (b_1, b_2)^T$.

By Taylor expansion, we have

$$\begin{aligned} 0 &= Q_{jn}(\alpha_0, 0, 0) \\ &\quad + \frac{\partial}{\partial \alpha} Q_{jn}(\alpha^*, b_1^*, b_2^*)(\hat{\alpha} - \alpha_0) + \frac{\partial}{\partial b_1} Q_{jn}(\alpha^*, b_1^*, b_2^*)\hat{b}_1 + \frac{\partial}{\partial b_2} Q_{jn}(\alpha^*, b_1^*, b_2^*)\hat{b}_2, \end{aligned} \tag{4.18}$$

where α^* lies between $\hat{\alpha}$ and α_0, and $b^* = (b_1^*, b_2^*)^T$ lies between $\hat{b} = (\hat{b}_1, \hat{b}_2)^T$ and 0.

Using Lemma 4.3, (4.16) and the fact that $\max_{1 \le i \le n} \gamma_i = \max_{1 \le i \le n} |b^T V_i(\alpha)| =$

$o_p(1)$, it is straightforward to check that, with probability 1,

$$\begin{aligned}
\lim_{n\to\infty}\frac{\partial}{\partial\alpha}Q_{1,2n}(\alpha,b_1,b_2) &= \lim_{n\to\infty}\frac{1}{nh}\sum_{i=1}^{n}\{\frac{V_i'(\alpha)}{1+b^TV_i(\alpha)}-\frac{V_i(\alpha)b^TV_i'(\alpha)}{\{1+b^TV_i(\alpha)\}^2}\}\\
&= \lim_{n\to\infty}\frac{1}{nh}\sum_{i=1}^{n}V_i'(\alpha) = \lim_{n\to\infty}\frac{1}{nh}\sum_{i=1}^{n}(-W_i, 2W_i\{\frac{\theta_0-\alpha}{\hat{Q}_\epsilon^2(p)}-\alpha\})\\
&= (a_1,a_2)^T,\\
\lim_{n\to\infty}\frac{\partial}{\partial b^T}Q_{1,2n}(\alpha,b_1,b_2) &= \lim_{n\to\infty}\{-\frac{1}{nh}\sum_{i=1}^{n}\frac{V_i(\alpha)V_i^T(\alpha)}{\{1+b^TV_i(\alpha)\}^2}\} = -S_0,\\
\lim_{n\to\infty}\frac{\partial}{\partial\alpha}Q_{3n}(\alpha,b_1,b_2) &= \lim_{n\to\infty}\frac{1}{nh}\sum_{i=1}^{n}\{\frac{b^TV_i''(\alpha)}{1+b^TV_i(\alpha)}-\frac{\{b^TV_i'(\alpha)\}^2}{\{1+b^TV_i(\alpha)\}^2}\} = 0,\\
\lim_{n\to\infty}\frac{\partial}{\partial b}Q_{3n}(\alpha,b_1,b_2) &= \lim_{n\to\infty}\frac{\partial}{\partial\alpha}Q_{1,2n}(\alpha,b_1,b_2) = (a_1,a_2)^T,
\end{aligned}$$

where $a_1 = -f^2(x)\tau_2$, $a_2 = 2\{\frac{\sigma(x)}{Q_\epsilon(p)} - m(x)\}f^2(x)\tau_2$ and all the above limits hold uniformly in $\alpha \in \{\alpha : |\alpha-\alpha_0| \le n^{-1/2+\delta}\}$. Hence, by solving (7.30), we have

$$\begin{pmatrix}\hat{b}\\ \hat{\alpha}-\alpha_0\end{pmatrix} = -M_n^{-1}(\alpha^*,b_1^*,b_2^*)\begin{pmatrix}Q_{1,2n}(\alpha_0,0,0)\\ 0\end{pmatrix} = -M^{-1}\begin{pmatrix}Q_{1,2n}(\alpha_0,0,0)\\ 0\end{pmatrix}+o_p(1),$$

where

$$M_n(\alpha,b_1,b_2) = \begin{pmatrix}\frac{\partial Q_{1,2n}(\alpha,b_1,b_2)}{\partial b^T} & \frac{\partial Q_{1,2n}(\alpha,b_1,b_2)}{\partial\alpha}\\ \frac{\partial Q_{3n}(\alpha,b_1,b_2)}{\partial b^T} & \frac{\partial Q_{3n}(\alpha,b_1,b_2)}{\partial\alpha}\end{pmatrix} \to \begin{pmatrix}-S_0 & A\\ A^T & 0\end{pmatrix} := M = \begin{pmatrix}M_{11} & M_{12}\\ M_{21} & M_{22}\end{pmatrix}$$

and $A = (a_1,a_2)^T$. Therefore,

$$\begin{aligned}
& l_n(\theta_0,\hat{\alpha})\\
=\ & 2\sum_{i=1}^{n}\hat{b}^TV_i(\hat{\alpha}) - \sum_{i=1}^{n}\{\hat{b}^TV_i(\hat{\alpha})\}^2 + o_p(1)\\
=\ & -nhQ_{1,2n}^T(\alpha_0,0,0)M_{11}^{-1}\{I+M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1}\}Q_{1,2n}(\alpha_0,0,0)+o_p(1)\\
=\ & \{(-M_{11})^{-\frac{1}{2}}\sqrt{nh}Q_{1,2n}(\alpha_0,0,0)\}^T\{I-(-M_{11})^{-\frac{1}{2}}M_{12}M_{22.1}^{-1}M_{21}(-M_{11})^{-\frac{1}{2}}\}\\
& \times\{(-M_{11})^{-\frac{1}{2}}\sqrt{nh}Q_{1,2n}(\alpha_0,0,0)\}+o_p(1).
\end{aligned}$$

Since $(-M_{11})^{-\frac{1}{2}}\sqrt{nh}Q_{1,2n}(\alpha_0,0,0)$ converges in distribution to a standard bivariate normal distribution and $I-(-M_{11})^{-\frac{1}{2}}M_{12}M_{22.1}^{-1}M_{21}(-M_{11})^{-\frac{1}{2}}$ is symmetric and idempotent with rank 1, we have

$$l_n(\theta_0, \hat{\alpha}) \xrightarrow{d} \chi_1^2,$$

which completes the proof. $\square$

Table 4.1: Coverage probabilities for model (i).

(n, h, x)	I_γ		I_γ^C		I_γ^N	
	γ=0.90	0.95	0.90	0.95	0.90	0.95
$(200, 0.5n^{-7/24}, 0.25)$	0.804	0.865	0.863	0.913	0.820	0.876
$(200, 0.5n^{-7/24}, 0.5)$	0.830	0.880	0.879	0.915	0.843	0.890
$(200, n^{-7/24}, 0.25)$	0.815	0.883	0.854	0.918	0.815	0.874
$(200, n^{-7/24}, 0.5)$	0.839	0.891	0.867	0.931	0.849	0.892
$(1000, 0.5n^{-7/24}, 0.25)$	0.875	0.933	0.905	0.960	0.872	0.932
$(1000, 0.5n^{-7/24}, 0.5)$	0.847	0.920	0.888	0.943	0.858	0.918
$(1000, n^{-7/24}, 0.25)$	0.858	0.913	0.878	0.939	0.859	0.908
$(1000, n^{-7/24}, 0.5)$	0.862	0.931	0.886	0.948	0.863	0.927

Table 4.2: Coverage probabilities for model (ii).

(n, h, x)	I_γ		I_γ^C		I_γ^N	
	γ=0.90	0.95	0.90	0.95	0.90	0.95
$(200, 0.5n^{-7/24}, 0.25)$	0.798	0.864	0.855	0.913	0.817	0.873
$(200, 0.5n^{-7/24}, 0.5)$	0.830	0.874	0.876	0.914	0.841	0.886
$(200, n^{-7/24}, 0.25)$	0.825	0.901	0.887	0.931	0.832	0.885
$(200, n^{-7/24}, 0.5)$	0.839	0.892	0.869	0.927	0.843	0.884
$(1000, 0.5n^{-7/24}, 0.25)$	0.873	0.935	0.899	0.959	0.870	0.929
$(1000, 0.5n^{-7/24}, 0.5)$	0.849	0.920	0.891	0.948	0.856	0.916
$(1000, n^{-7/24}, 0.25)$	0.851	0.914	0.881	0.939	0.852	0.906
$(1000, n^{-7/24}, 0.5)$	0.863	0.923	0.880	0.944	0.859	0.917

CHAPTER V

INTERMEDIATE QUANTILES

Intermediate quantiles play an important role in the statistics of extremes with particular applications in risk management. For interval estimation of quantiles, Chen and Hall (1993) proposed the so-called smoothed empirical likelihood method. In this chapter, we apply the method in Chen and Hall (1993) to construct confidence intervals for an intermediate quantile by deriving the corresponding Wilks Theorem. The content of this chapter is based on Z. Li, Y. Gong, and L. Peng (2010), Empirical likelihood methods for intermediate Quantiles, *Statistics and Probability Letters 80, 1022–1029*.

5.1 Introduction

Suppose $X_1, \cdots, X_n$ are independent and identically distributed random variables with distribution function F. The q-th quantile of F is defined as $F^{-}(q)$, where F^{-} denotes the inverse function of F. Quantiles are of importance in statistical inference, and both empirical quantile estimation and kernel smooth quantile estimation have been studied in the literature for a long history. Interval estimation for a quantile includes the normal approximation method, the bootstrap method, the jackknife method and the empirical likelihood method.

When one applies Owen's method to quantiles directly, the resultant interval has exact coverage probability $P(r_1 \leq M < r_2)$ for some integers $r_1 < r_2$, where M is a binomial random variable. Because of the discreteness of the binomial distribution, the set of possible confidence levels is rather sparse even for moderate sample size. Chen and Hall (1993) proposed the so-called smoothed empirical likelihood method to construct confidence intervals for the q-th quantile and showed that smoothing is necessary to achieve Bartlett correction. Smoothing also makes the optimization in the empirical likelihood method easy in genenal. Another way of smoothing the empirical likelihood ratio was proposed by Adimari (1998).

Given the advantages of the smoothed empirical likelihood method for quantiles, in this paper we investigate the feasibility of applying the method in Chen and Hall (1993) to an intermediate quantile. When $q = q_n \to 1$ and $n(1-q_n) \to \infty$ as $n \to \infty$, we call $F^{-}(q_n)$ an intermediate quantile. Intermediate quantiles play an important role in the statistics of extremes with particular applications to risk management. For example, Pickands (1975) used three intermediate order statistics to estimate the extreme value index; Viharos (1997) employed a linear combination of intermediate order statistics to estimate the tail index of a heavy tailed distribution; Csörgő and Steinebach (1991) applied intermediate order statistics to estimate the adjustment coefficient in risk theory. More references on the study of intermediate quantiles can be found in Peng and Yang (2009).

Because the intermediate quantile $F^{-}(q_n)$ tends to the right endpoint of the underlying distribution function F, some conditions on the tail behavior of F are needed in order to derive the asymptotic limit of the empirical intermediate quantile. Extreme value conditions are employed for such a study; see Dekkers and de Haan (1989), Dekkers, Einmahl and de Haan (1989) and Csörgő and Horváth (1993). In this chapter, using extreme value conditions, we show that the method in Chen and Hall (1993) is applicable to an intermediate quantile, but the choice of bandwidth depends on the tail behavior.

5.2 Methodologies and main results

Throughout the chapter, we assume that $K(x)$ is a symmetric density with support in $[-1, 1]$. Put $G(x) = \int_{-\infty}^{x} K(y)dy$. As in Chen and Hall (1993), we define the smoothed empirical likelihood for the q_n-th intermediate quantile $\theta_n = F^{-}(q_n)$ as

$$L_n(q_n, \theta_n) = \sup\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i G(\frac{\theta_n - X_i}{h}) = q_n, p_i > 0, i = 1, \ldots, n\},$$

where $h = h(n) > 0$ is a bandwidth.

Define $w_i = w_i(\theta_n) = G((\theta_n - X_i)/h) - q_n$. Since $\prod_{i=1}^{n} p_i$ attains its maximum at $p_i = 1/n$, the empirical likelihood ratio at θ_n is defined as

$$R_n(q_n, \theta_n) = \sup\{\prod_{i=1}^{n}(np_i) : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i w_i = 0, p_i > 0, i = 1, \ldots, n\}.$$

By applying the standard method of Lagrange multiplier, we know that $R_n(q_n, \theta_n)$ attains its maximum at

$$p_i = \frac{1}{n}\frac{1}{1 + \lambda w_i},$$

where $\lambda = \lambda(\theta_n)$ is determined by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{w_i}{1 + \lambda w_i} = 0. \tag{5.1}$$

Thus, the empirical log likelihood ratio is given by

$$l_n(q_n, \theta_n) = -2\log R_n(q_n, \theta_n) = 2\sum_{i=1}^{n}\log(1 + \lambda w_i). \tag{5.2}$$

In order to derive the asymptotic properties of $l_n(q_n, \theta_n)$, we employ the following von Mises' condition:

$$\lim_{t \to \theta^*}\frac{\{1 - F(t)\}F''(t)}{\{F'(t)\}^2} = -\gamma - 1 \tag{5.3}$$

for some $\gamma \in R$, where $\theta^* = \sup\{x : F(x) < 1\}$. Note that the above condition implies that F lies in the domain of attraction of an extreme value distribution with index γ; see Theorem 1.1.8 of de Haan and Ferreira (2006).

Theorem 5.1. *(Li, Gong and Peng, 2010) Assume that $q_n \to 1$ and $n(1-q_n) \to \infty$ as $n \to \infty$. Suppose $F''(x)$ exists and $F'(x)$ is positive for all x in some left neighborhood of θ^*. If (5.3) holds for some $\gamma \neq 0$ and*

$$\begin{cases} n(1 - q_n)\{\frac{h}{F^{-}(q_n)}\}^4 \to 0, & \textit{when} \quad \gamma > 0, \\ n(1 - q_n)\{\frac{h}{\theta^* - F^{-}(q_n)}\}^4 \to 0, & \textit{when} \quad \gamma < 0, \end{cases} \tag{5.4}$$

as $n \to \infty$*, then* $l_n(q_n, \theta_n) \xrightarrow{\text{d}} \chi_1^2$ *as* $n \to \infty$*.*

Based on the above theorem, an empirical likelihood based confidence interval for θ_n with level α can be obtained as

$$I_\alpha(h, n) = \{\beta_n : l_n(q_n, \beta_n) \le z_\alpha\},$$

where z_α is chosen to satisfy $P(\chi_1^2 \le z_\alpha) = \alpha$.

When $\gamma < 0$, condition (5.4) implies that $h \to 0$ since $n(1 - q_n) \to \infty$ implies that $h/\{\theta^* - F^-(q_n)\} \to 0$ as $n \to \infty$. However, when $\gamma > 0$, condition (5.4) does not imply that the bandwidth h has to tend to zero. Here we propose the following modified empirical likelihood method, which requires that the bandwidth tends to zero.

Define $w_i^* = w_i^*(\theta_n) = G((1 - X_i/\theta_n)/h) - q_n$ for $i = 1, \cdots, n$ and the empirical likelihood function at θ_n as

$$R_n^*(q_n, \theta_n) = \sup\{\prod_{i=1}^n (np_i) : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i w_i^* = 0, p_1 \ge 0, \cdots, p_n \ge 0\}.$$

Then the convergence of the empirical log likelihood ratio is derived in the following theorem.

Theorem 5.2. *(Li, Gong and Peng, 2010) Assume that* $q_n \to 1$ *and* $n(1 - q_n) \to \infty$ *as* $n \to \infty$*. Suppose* $F''(x)$ *exists, (5.3) holds for some* $\gamma > 0$ *and*

$$n(1 - q_n)h^4 \to 0 \quad \text{as} \quad n \to \infty, \tag{5.5}$$

then $-2 \log R_n^*(q_n, \theta_n) \xrightarrow{\text{d}} \chi_1^2$ *as* $n \to \infty$*.*

Hence, in case of $\gamma > 0$, an empirical likelihood based confidence interval for θ_n with level α can be obtained as

$$I_\alpha^*(h, n) = \{\beta_n : -2 \log R_n^*(q_n, \beta_n) \le z_\alpha\},$$

where z_α is the α quantile of χ_1^2.

Remark 5.1. *Note that (5.3) implies that, for all $x > 0$,*

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma} \quad \textit{when} \quad \gamma > 0, \quad \textit{and} \quad \lim_{t\to\infty} \frac{U(\infty) - U(tx)}{U(\infty) - U(t)} = x^{\gamma} \quad \textit{when} \quad \gamma < 0.$$

Hence, as n large enough, we have, for any $\epsilon > 0$,

$$F^{-}(q_n) = U(\frac{1}{1-q_n}) \geq (\frac{1}{1-q_n})^{\gamma-\epsilon} \quad \textit{when} \quad \gamma > 0,$$

and

$$\theta^* - F^{-}(q_n) = U(\infty) - U(\frac{1}{1-q_n}) \geq (\frac{1}{1-q_n})^{\gamma-\epsilon} \quad \textit{when} \quad \gamma < 0.$$

Therefore, a sufficient condition for (5.4) is

$$\lim_{n\to\infty} nh^4(1-q_n)^{4\gamma-4\epsilon+1} = 0 \tag{5.6}$$

for some $\epsilon > 0$. Condition (5.6) can be employed to choose the bandwidth via estimating the extreme value index γ.

Remark 5.2. *Let $\hat{\theta}_n = \arg\max_{\theta_n} R_n(q_n, \theta_n)$ denote the maximum empirical likelihood estimate. Then by the standard arguments in Qin and Lawless (1994) and Lemma 5.1 below, $\hat{\theta}_n - \theta_n = O_p(\{n(1-q_n)\}^{-1/2})$ under conditions of Theorem 5.1. The same rate of convergence holds for $\hat{\theta}_n^* = \arg\max_{\theta_n} R_n^*(q_n, \theta_n)$ under the conditions of Theorem 5.2.*

5.3 Simulation study

In this section, we investigate the finite sample behavior of the proposed empirical likelihood methods in terms of coverage accuracy by drawing $10,000$ random samples from the extreme value distribution $G(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\}$, where $1 + \gamma x > 0$ and $\gamma \in R$.

First we consider the case of $\gamma > 0$, i.e., heavy tailed distributions which have been employed in risk management. For computing $I_\alpha^*(h, n)$, we use the kernel $k(x) = \frac{15}{16}(1-x^2)^2 I(|x| \leq 1)$, $h = c\{n(1-q_n)\}^{-1/3}$ and $q_n = 1 - n^{-a}$. The choice of h is motivated by that in Chen and Hall (1993). In Table 5.1, we report the coverage probability errors for the cases of $n = 1000, 10,000$, $c = 0.1, 0.5$ and $a = 0.5, 0.8$. In comparison with Tables 1 and 4 in Peng and Yang (2009), we observe that the proposed empirical likelihood method performs better than the jackknife method in Peng and Yang (2009) when $\gamma = 5$, but slightly worse when $\gamma = 0.2$.

Next we consider the case of $\gamma \neq 0$, i.e., the confidence intervals $I_\alpha(h, n)$. We employ the same kernel and q_n as above, but choose $h = 0.001n^{-1/4}(1 - q_n)^{-\hat{\gamma}(k)-1/4}$ for the case of $\gamma = \pm 5$ and $h = 0.1n^{-1/4}(1 - q_n)^{-\hat{\gamma}(k)-1/4}$ for the case of $\gamma = \pm 0.2$, where $\hat{\gamma}(k)$ is the estimator for γ given in Dekkers, Einmahl and de Haan (1989) and k is the number of upper order statistics involved in the estimation. In Table 5.2, we report the coverage probability errors for the cases $n = 1000$, $a = 0.5, 0.8$, and $k = 50, 100, 200$. This table shows that the choice of k, i.e., the estimator for γ plays an important role. In comparison with Tables 1 and 4 in Peng and Yang (2009), the proposed empirical likelihood method performs better than the jackknife method in Peng and Yang (2009) when $a = 0.5$, but performs worse when $a = 0.8$.

5.4 Proofs

Throughout we define $\bar{w}_j = \frac{1}{n}\sum_{i=1}^n w_i^j$ and $\mu_j = E(\bar{w}_j)$ for $j = 1, 2$. First we show the following lemma.

Lemma 5.1. *Under conditions of Theorem 1, we have, as $n \to \infty$,*

$$\begin{cases} \mu_1 = \frac{h^2}{2}F''(\theta_n)(1 + o(1))\int_{-1}^1 z^2K(z)dz, & \text{when} \quad \gamma \neq -1, \\ \mu_1 = o(\frac{h^2(1-q_n)}{\{\theta^* - F^-(q_n)\}^2}), & \text{when} \quad \gamma = -1, \\ \mu_2 = (1 - q_n)(1 + o(1)), & \\ \dfrac{n\mu_1^2}{\mu_2} \to 0. & \end{cases}$$

Proof. Let $U(t)$ denote the inverse of $1/\{1 - F(x)\}$. Then Corollary 1.1.10 of de Haan and Ferreira (2006) says that (5.3) implies that

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{tU'(t)} = \frac{x^\gamma - 1}{\gamma} \quad \text{for all} \quad x > 0. \tag{5.7}$$

Applying Theorem B.2.2 of de Haan and Ferreira (2006) to (5.7), we have

$$\begin{cases} \lim_{t\to\infty} \frac{tU'(t)}{U(t)} = \gamma, \quad \text{when} \quad \gamma > 0, \\ \lim_{t\to\infty} \frac{tU'(t)}{U(\infty)-U(t)} = -\gamma, \quad \text{when} \quad \gamma < 0. \end{cases} \tag{5.8}$$

By (5.8), the proofs of Theorems 1.1.11 and 1.1.13 of de Haan and Ferreira (2006), we have

$$\begin{cases} \lim_{t\to\infty} \frac{tF'(t)}{1-F(t)} = \gamma^{-1}, \quad \text{when} \quad \gamma > 0, \\ \lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma} \quad \text{for all} \quad x > 0, \quad \text{when} \quad \gamma > 0, \\ \lim_{t\to\theta^*} \frac{(\theta^*-t)F'(t)}{1-F(t)} = -\gamma^{-1}, \quad \text{when} \quad \gamma < 0, \\ \lim_{t\to 0} \frac{1-F(\theta^*-tx)}{1-F(\theta^*-t)} = x^{-1/\gamma}, \quad \text{for all} \quad x > 0, \quad \text{when} \quad \gamma < 0. \end{cases} \tag{5.9}$$

It follows from Taylor's expansion that

$$\begin{aligned} \mu_1 = E\bar{w}_1 &= E\{G(\frac{\theta_n - X_i}{h}) - q_n\} \\ &= \int_{-\infty}^{\infty} G(\frac{\theta_n - x}{h})dF(x) - q_n \\ &= \int_{-1}^{1} F(\theta_n - hz)K(z)dz - q_n \\ &= \int_{-1}^{1} \{F(\theta_n) - hzF'(\theta_n) + \frac{h^2z^2}{2}F''(\theta_n^*(z))\}K(z)dz - q_n \\ &= \frac{h^2}{2}\int_{-1}^{1} z^2K(z)F''(\theta_n^*(z))dz, \end{aligned} \tag{5.10}$$

where $\theta_n^*(z) \in (\theta_n - h, \theta_n + h)$ for all $z \in [-1, 1]$. Note that conditions in Theorem 5.1 imply that

$$\begin{cases} \lim_{n\to\infty} \frac{h}{F^-(q_n)} = 0 \quad \text{when} \quad \gamma > 0, \\ \lim_{n\to\infty} \frac{h}{\theta^*-F^-(q_n)} = 0 \quad \text{when} \quad \gamma < 0, \end{cases} \tag{5.11}$$

which implies that

$$\begin{cases} \lim_{n\to\infty} \frac{\theta_n^*(z)}{\theta_n} = 1 \quad \text{uniformly in} \quad z \in [-1, 1] \quad \text{when} \quad \gamma > 0, \\ \lim_{n\to\infty} \frac{\theta^*-\theta_n^*(z)}{\theta^*-\theta_n} = 1 \quad \text{uniformly in} \quad z \in [-1, 1] \quad \text{when} \quad \gamma < 0. \end{cases} \tag{5.12}$$

By (5.3), (5.9) and (5.12), we have

$$\begin{cases} \lim_{n\to\infty} F''(\theta_n^*(z))/F''(\theta_n) = 1 \quad \text{uniformly in} \quad z \in [-1,1] \quad \text{when} \quad \gamma \neq -1, \\ F''(\theta_n^*(z)) = o(\frac{1-q_n}{\{\theta^*-F^-(q_n)\}^2}) \quad \text{uniformly in} \quad z \in [-1,1] \quad \text{when} \quad \gamma = -1, \\ F''(\theta_n) \sim \frac{-\gamma-1}{\gamma^2}\frac{1-q_n}{\{F^-(q_n)\}^2} \quad \text{when} \quad \gamma > 0, \\ F''(\theta_n) \sim \frac{-\gamma-1}{\gamma^2}\frac{1-q_n}{\{\theta^*-F^-(q_n)\}^2} \quad \text{when} \quad \gamma < 0 \quad \text{and} \quad \gamma \neq -1, \\ F''(\theta_n) = o(\frac{\{F'(\theta_n)\}^2}{1-F(\theta_n)}) \quad \text{when} \quad \gamma = -1. \end{cases} \tag{5.13}$$

It follows from (5.10) and (5.13) that

$$\begin{cases} \mu_1 = \frac{h^2}{2}F''(\theta_n)(1+o(1))\int_{-1}^1 z^2K(z)dz \quad \text{when} \quad \gamma \neq -1, \\ \mu_1 = o(\frac{h^2(1-q_n)}{\{\theta^*-F^-(q_n)\}^2}) \quad \text{when} \quad \gamma = -1. \end{cases} \tag{5.14}$$

Write

$$\begin{aligned} \mu_2 = E\bar{w}_2 &= E\left\{G\left(\tfrac{\theta_n-X_1}{h}\right) - q_n\right\}^2 \\ &= E\left\{G^2\left(\tfrac{\theta_n-X_1}{h}\right)\right\} - 2q_nE\left\{G\left(\tfrac{\theta_n-X_1}{h}\right)\right\} + q_n^2. \end{aligned} \tag{5.15}$$

By Taylor's expansion, we have

$$\begin{aligned} & E\left\{G^2\left(\tfrac{\theta_n-X_1}{h}\right)\right\} \\ =\ & \int_{-\infty}^{\infty} G^2\left(\tfrac{\theta_n-x}{h}\right)dF(x) \\ =\ & \tfrac{2}{h}\int_{-\infty}^{\infty} F(x)G\left(\tfrac{\theta_n-x}{h}\right)K\left(\tfrac{\theta_n-x}{h}\right)dx \\ =\ & 2\int_{-1}^1 F(\theta_n-hz)G(z)K(z)dz \\ =\ & 2\int_{-1}^1 \{F(\theta_n) - hzF'(\theta_n) + \tfrac{h^2z^2}{2}F''(\theta_n^*(z))\}K(z)G(z)dz \\ =\ & F(\theta_n) - 2hF'(\theta_n)\int_{-1}^1 zK(z)G(z)dz + h^2\int_{-1}^1 z^2F''(\theta_n^*(z))K(z)G(z)dz \\ =\ & q_n - 2hF'(\theta_n)\int_{-1}^1 zK(z)G(z)dz + h^2\int_{-1}^1 z^2F''(\theta_n^*(z))K(z)G(z)dz. \end{aligned} \tag{5.16}$$

Using (5.13), (5.14), (5.15) and (5.16), we have

$$\begin{aligned} \mu_2 &= q_n - q_n^2 - 2hF'(\theta_n)\int_{-1}^1 zK(z)G(z)dz \\ &\quad + h^2F''(\theta_n)(1+o(1))\{\int_{-1}^1 z^2K(z)G(z)dz - q_n\int_{-1}^1 z^2K(z)\,dz\} \end{aligned} \tag{5.17}$$

when $\gamma \neq -1$, and

$$\mu_2 = q_n - q_n^2 - 2hF'(\theta_n)\int_{-1}^1 zK(z)G(z)dz + o(\frac{h^2(1-q_n)}{\{\theta^* - F^-(q_n)\}^2}) \tag{5.18}$$

when $\gamma = -1$. By (5.9), (5.11) and (5.13), we have

$$\begin{cases} \frac{hF'(\theta_n)}{1-q_n} \sim \gamma^{-1}\frac{h}{F^-(q_n)} = o(1) \quad \text{when} \quad \gamma > 0, \\ \frac{hF'(\theta_n)}{1-q_n} \sim -\gamma^{-1}\frac{h}{\theta^*-F^-(q_n)} = o(1) \quad \text{when} \quad \gamma < 0, \\ \frac{h^2F''(\theta_n)}{1-q_n} \sim \frac{-\gamma-1}{\gamma^2}\{\frac{h}{F^-(q_n)}\}^2 = o(1) \quad \text{when} \quad \gamma > 0, \\ \frac{h^2F''(\theta_n)}{1-q_n} \sim \frac{-\gamma-1}{\gamma^2}\{\frac{h}{\theta^*-F^-(q_n)}\}^2 = o(1) \quad \text{when} \quad \gamma < 0 \quad \text{and} \quad \gamma \neq -1, \\ \frac{h^2F''(\theta_n)}{1-q_n} = o(\{\frac{h}{\theta^*-F^-(q_n)}\}^2) = o(1) \quad \text{when} \quad \gamma = -1. \end{cases} \tag{5.19}$$

It follows from (5.13), (5.14), (5.17), (5.18), (5.19) and conditions in Theorem 5.1 that

$$\begin{cases} \mu_2 = (1-q_n)(1+o(1)), \\ \frac{n\mu_1^2}{\mu_2} = o(1). \end{cases} \tag{5.20}$$

Hence, the lemma follows from (5.14) and (5.20).

□

Proof of Theorem 5.1. Observe that

$$\begin{aligned} 0 &= n^{-1}|\sum_{i=1}^n w_i(1+\lambda w_i)^{-1}| \\ &= n^{-1}|\sum_{i=1}^n \{\lambda w_i^2(1+\lambda w_i)^{-1} - w_i\}| \\ &\geq |\lambda|(1+|\lambda|)^{-1}\bar{w}_2 - |\bar{w}_1|, \end{aligned}$$

since $\max_{1\leq i\leq n}|w_i| \leq 1$. Thus,

$$|\lambda|(1 - \frac{|\bar{w}_1|}{\bar{w}_2}) \leq \frac{|\bar{w}_1|}{\bar{w}_2}. \tag{5.21}$$

Since $E|w_i|^3 \leq Ew_i^2 = \mu_2$, it follows from Lemma 5.1 that

$$\sum_{i=1}^n E|w_i|^3 \leq \sum_{i=1}^n Ew_i^2 = n\mu_2 = o(\{n(1-q_n)\}^{3/2}) = o(\{\sum_{i=1}^n E(w_i-\mu_1)^2\}^{3/2}).$$

Hence, by the central limit theorem (see Corollary 1, page 298 of Chow and Teicher (1988)), we have

$$\frac{\sqrt{n}\{\bar{w}_1 - \mu_1\}}{\sqrt{\mu_2 - \mu_1^2}} \xrightarrow{d} N(0,1). \tag{5.22}$$

Since $|w_i| \leq 1$, we have

$$P(|\bar{w}_2 - \mu_2| \geq \epsilon) \leq \frac{\sum_{i=1}^n E\{w_i^2 - \mu_2\}^2}{n^2\epsilon^2} \leq \frac{1}{n\epsilon^2}$$

for any $\epsilon > 0$, i.e.,

$$\bar{w}_2 \xrightarrow{p} \mu_2. \tag{5.23}$$

It follows from Lemma 5.1, (6.2) and (5.23) that

$$\begin{aligned}
\frac{\bar{w}_1}{\bar{w}_2} &= \frac{\frac{\sqrt{n}(\bar{w}_1-\mu_1)}{\sqrt{\mu_2-\mu_1^2}} + \frac{\sqrt{n}\mu_1}{\sqrt{\mu_2-\mu_1^2}}}{\frac{\sqrt{n}\mu_2}{\sqrt{\mu_2-\mu_1^2}}\frac{\bar{w}_2}{\mu_2}} \\
&= O_p(\frac{\sqrt{\mu_2-\mu_1^2}}{\sqrt{n}\mu_2}) \\
&= O_p(\frac{1}{\sqrt{n(1-q_n)}}).
\end{aligned}$$

Hence, (5.21) implies that

$$\lambda = O_p(\frac{1}{\sqrt{n(1-q_n)}}). \tag{5.24}$$

Now

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^n \frac{w_i}{1+\lambda w_i} \\
&= \frac{1}{n}\sum_{i=1}^n w_i\{1 - \lambda w_i + \frac{(\lambda w_i)^2}{1+\lambda w_i}\} \\
&= \bar{w}_1 - \lambda\bar{w}_2 + n^{-1}\sum_{i=1}^n \frac{\lambda^2 w_i^3}{1+\lambda w_i} \\
&= \bar{w}_1 - \lambda\bar{w}_2 + O_p(\lambda^2\bar{w}_2).
\end{aligned}$$

Thus,

$$\lambda = \bar{w}_1\bar{w}_2^{-1} + O_p(\frac{1}{n(1-q_n)}). \tag{5.25}$$

Applying Taylor's expansion to (5.2) and using (5.25), we can show that

$$\begin{aligned}
l_n(q_n, \theta_n) &= 2\sum_{i=1}^{n} \log(1+\lambda w_i) \\
&= 2n\lambda \bar{w}_1 - n\lambda^2 \bar{w}_2 + 2\sum_{i=1}^{n} \eta_i \\
&= n\bar{w}_1^2 \bar{w}_2^{-1} - n\bar{w}_2 O_p\left(\frac{1}{n^2(1-q_n)^2}\right) + 2\sum_{i=1}^{n} \eta_i \\
&= \frac{\left\{\frac{\sqrt{n}(\bar{w}_1-\mu_1)}{\sqrt{\mu_2-\mu_1^2}} + \frac{\sqrt{n}\mu_1}{\sqrt{\mu_2-\mu_1^2}}\right\}^2}{\frac{\mu_2}{\mu_2-\mu_1^2}\frac{\bar{w}_2}{\mu_2}} - n\bar{w}_2 O_p\left(\frac{1}{n^2(1-q_n)^2}\right) + 2\sum_{i=1}^{n} \eta_i, \quad (5.26)
\end{aligned}$$

where

$$P(|\eta_i| \leq C|\lambda w_i|^3, 1 \leq i \leq n) \to 1 \quad \text{as} \quad n \to \infty$$

for some constant $C > 0$. By (5.24) and Lemma 5.1, we have

$$2\sum_{i=1}^{n} \eta_i = O_p(|\lambda|^3 \sum_{i=1}^{n} w_i^2) = O_p\left(\frac{1}{\sqrt{n(1-q_n)}}\right). \quad (5.27)$$

Hence, Theorem 5.1 follows from Lemma 5.1, (5.24), (5.26), (5.27) and (6.2). $\square$

Proof of Theorem 5.2. It can be proved in a similar way to the proof of Theorem 5.2. $\square$

Table 5.1: The absolute coverage probability errors for the confidence intervals $I^*_{0.9}(h,n)$ and $I^*_{0.95}(h,n)$ are reported for $q_n = 1 - n^{-a}$ and $h = c\{n(1-q_n)\}^{-1/3}$.

(γ, n, a)	$I^*_{0.9}$ $c = 0.1$	$I^*_{0.9}$ $c = 0.5$	$I^*_{0.95}$ $c = 0.1$	$I^*_{0.95}$ $c = 0.5$
(0.2, 1000,0.5)	0.0021	0.0032	0.0030	0.0025
(0.2, 10000,0.5)	0.0017	0.0032	0.0008	0.0020
(0.2, 1000,0.8)	0.0311	0.0238	0.0025	0.0194
(0.2, 10000,0.8)	0.0075	0.0173	0.0102	0.0122
(5, 1000,0.5)	0.0116	0.0029	0.0005	0.0020
(5, 10000,0.5)	0.0026	0.0017	0.0011	0.0004
(5, 1000,0.8)	0.0449	0.0406	0.0115	0.0097
(5, 10000,0.8)	0.0068	0.0057	0.0281	0.0217

Table 5.2: The absolute coverage probability errors for the confidence intervals $I_{0.9}(h,n)$ and $I_{0.95}(h,n)$ are reported for $q_n = 1 - n^{-a}$, $n = 1000$ and $h = cn^{-1/4}(1-q_n)^{-\hat{\gamma}(k)-1/4}$.

(γ, a, c)	$I_{0.9}$ $k = 50$	$I_{0.9}$ $k = 100$	$I_{0.9}$ $k = 200$	$I_{0.95}$ $k = 50$	$I_{0.95}$ $k = 100$	$I_{0.95}$ $k = 200$
(-5, 0.5,0.001)	0.0240	0.0062	0.0125	0.0306	0.0064	0.0006
(-5, 0.8,0.001)	0.0696	0.0474	0.0459	0.0204	0.0003	0.0104
(-0.2, 0.5,0.1)	0.0025	0.0016	0.0018	0.0008	0.0020	0.0025
(-0.2, 0.8,0.1)	0.0407	0.0409	0.0411	0.0093	0.0095	0.0092
(0.2, 0.5,0.1)	0.0009	0.0010	0.0004	0.0017	0.0019	0.0029
(0.2, 0.8,0.1)	0.0418	0.0422	0.0408	0.0097	0.0088	0.0079
(5, 0.5,0.001)	0.0014	0.0103	0.0119	0.0108	0.0021	0.0007
(5, 0.8,0.001)	0.0992	0.0675	0.0550	0.0545	0.0180	0.0017

CHAPTER VI

COVERAGE ACCURACY FOR A MEAN WITHOUT THIRD MOMENT

For constructing a confidence interval for the mean of a random variable with a known variance, one may prefer the sample mean standardized by the true standard deviation to the Student's t-statistic since the information of knowing the variance is used in the former way. In this chapter, by comparing the leading error term in the expansion of the coverage probability, we show that the above statement is not true when the third moment is infinite. Our theory prefers the Student's t-statistic either when one-sided confidence intervals are considered for a heavier tail distribution or when two-sided confidence intervals are considered. Unlike other existing expansions for the Student's t-statistic, the derived explicit expansion for the case of infinite third moment can be used to estimate the coverage error so that bias correction becomes possible. The content of this chapter is based on Y. Gong and L. Peng (2010), Coverage accuracy for a mean without third moment, *Journal of Statistical Planning and Inference, 104, 1082–1088*.

6.1 Introduction

Let $X, X_1, \cdots, X_n$ be independent and identically distributed random variables with distribution function F, which lies in the domain of attraction of a normal law, i.e.,

$$\lim_{t\to\infty} \frac{\int_0^{tx}\{1-F(u)+F(-u)\}u\,du}{\int_0^{t}\{1-F(u)+F(-u)\}u\,du} = 1 \quad \text{for} \quad x>0. \tag{6.1}$$

Put $\mu = E(X)$. Then there exists constant $a_n > 0$ such that

$$\lim_{n\to\infty} P(a_n(\bar{X}_n-\mu)\le x) = \Phi(x) := (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-y^2/2}\,dy \tag{6.2}$$

for all $x \in R$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, see Feller (1971). It is also true that condition (6.1) implies that the Student's t-statistic converges in distribution to a standard normal distribution, i.e.,

$$\lim_{n\to\infty} P(T_n \le x) = \Phi(x) \quad \text{for all} \quad x \in R, \tag{6.3}$$

where T_n is the so-called Student's t-statistic defined as

$$T_n = \sum_{i=1}^n (X_i - \mu)/\{\sum_{i=1}^n X_i^2 - n^{-1}(\sum_{i=1}^n X_i)^2\}^{1/2},$$

see Giné, Götze and Mason (1997).

When the third moment of X is finite and F is non-lattice, we have the following Edgeworth expansions

$$\lim_{n\to\infty} n^{1/2}\{P(\sqrt{n}(\bar{X}_n - \mu)/\sigma) \le x) - \Phi(x)\} = \frac{1-x^2}{6}\Phi'(x)\frac{E(X-\mu)^3}{\sigma^3} := l_1(x) \tag{6.4}$$

and

$$\lim_{n\to\infty} n^{1/2}\{P(T_n \le x) - \Phi(x)\} = \frac{2x^2+1}{6}\Phi'(x)\frac{E(X-\mu)^3}{\sigma^3} := l_2(x) \tag{6.5}$$

uniformly in $x \in R$, where $\sigma = \{E(X-\mu)^2\}^{1/2}$, see Hall (1990). Since $(1-x^2)^2 - (2x^2+1)^2 = -3x^4 - 6x^2 \le 0$, i.e., $l_1^2(x) - l_2^2(x) \le 0$ for all $x \in R$, (6.4) and (6.5) imply that one-sided confidence interval based on $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a more accurate coverage probability than that based on T_n when σ is known. This is not surprising at all since the former one employs the information that the variance is known. Also note that both $l_1(x)$ and $l_2(x)$ are even functions. Now the question is: do these two properties still hold when $EX^3 = \infty$?

To answer these questions, one has to derive expansions (6.4) and (6.5) without finite third moment. In this chapter, based on de Haan and Peng (1997) and Hall and Wang (2004), we derive the exact limits in (6.4) and (6.5) under the following conditions

$$\lim_{t\to\infty} \frac{1-G(tx)+G(-tx)}{1-G(t)+G(-t)} = x^{-2+\rho} \quad \text{and} \quad \lim_{t\to\infty} \frac{1-G(t)}{1-G(t)+G(-t)} = p \in [0,1], \tag{6.6}$$

for all $x > 0$ and some $\rho \in (-1, 0)$, where $G(x) = P(X - \mu \le x)$. Obviously, (6.6) with $\rho \in (-1, 0)$ implies $E|X|^3 = \infty$. Although these new expansions heavily depend on the known results in Hall and Wang (2004), these explicit expansions can be used to estimate the coverage error so that bias correction is possible. Bias correction is one of important motivations for the study of Edgeworth expansions. Since the expansions in Hall and Wang (2004) can not be estimated when the third moment is infinite (see section 6.2 below for details), the new results are important and practically useful. For the comparison study conducted in this paper, one may argue that knowing variance is not practical. However this comparison study gives a good example to show that including seemingly related information may not produce a better inference procedure. Moreover, the comparison study confirms the advantage of using the Student's t-statistic to deal with heavy tails. Let's summarize the main findings from the derived Edgeworth expansions in Theorem 6.1 below under conditions (6.6) as follows: i) neither of these two limits is an even function; ii) confidence intervals based on $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ are not always more accurate than those based on T_n in terms of coverage accuracy; iii) when ρ is near zero, i.e., F has heavier tails, or $p = 0.5$, one-sided confidence intervals based on the Student's t-statistic are preferred; iv) for constructing two-sided confidence intervals, the Student's t-statistic is preferred.

6.2 Main results

Under conditions (6.6), de Haan and Peng (1997) showed that

$$\begin{aligned} &\lim_{n\to\infty} \tfrac{P(\sum_{i=1}^n (X_i-\mu)/a_n \le x) - \Phi(x)}{n\{1-G(a_n)+G(-a_n)\}} \\ = &\tfrac{1}{2\pi}\int_{-\infty}^{\infty} t^{-1}\{-sgn(t)|t|^{2-\rho}A_\rho \cos(tx) + |t|^{2-\rho}B_\rho \sin(tx) + t^2 \tfrac{1-|t|^{-\rho}}{\rho}\sin(tx)\}e^{-t^2/2}\,dt \end{aligned} \tag{6.7}$$

uniformly in $x \in R$, where $a_n = a(n)$ with

$$a(x) = \sup\{a : 2a^{-2}H(a) \ge x^{-1}\}, \quad H(x) = \int_0^x \{1 - G(u) + G(-u)\}u\,du,$$

$$A_\rho = \frac{2p-1}{\rho(\rho-1)}\Gamma(1+\rho)\sin(\frac{\rho\pi}{2}), \quad B_\rho = \frac{1}{\rho(\rho-1)}\Gamma(1+\rho)\cos(\frac{\rho\pi}{2}) + \frac{1}{\rho},$$

and

$$sgn(t) = \begin{cases} 1 & t \geq 0 \\ -1 & t < 0. \end{cases}$$

Both rates of convergence and Edgeworth expansions for Student's t-statistics have been studied well in the literature, see Bentkus and Götze (1996), Hall (1987), Wang and Jing (1999) and Wang, Jing and Zhao (2000). Such studies can be employed not only to choose the sample size to ensure that the approximation error is under control, but also to improve the statistical inference by correcting the bias. Recently, under the minimal condition (6.1), Hall and Wang (2004) showed that

$$\sup_x |P(T_n \leq x) - \Phi(x) - L_n(x)| = o(\delta_n) + O(n^{-1/2}), \tag{6.8}$$

where

$$L_n(x) = nE\{\Phi(x\sqrt{1 + (\frac{X-\mu}{b_n})^2} - \frac{X-\mu}{b_n}) - \Phi(x)\},$$

$$b_n = \sup\{x : nx^{-2}E\{(X-\mu)^2 I(|X-\mu| \leq x)\} \geq 1\},$$

$$\begin{aligned} \delta_n &= nP(|X-\mu| > b_n) + nb_n^{-1}|E\{(X-\mu)I(|X-\mu| \leq b_n)\}| \\ &+ nb_n^{-3}|E\{(X-\mu)^3 I(|X-\mu| \leq b_n)\}| + nb_n^{-4}E\{(X-\mu)^4 I(|X-\mu| \leq b_n)\}. \end{aligned}$$

Estimating b_n can be done via replacing the expectation by the average. After estimating b_n, one may estimate $L_n(x)$ by using the same argument. However, from the proof of Theorem 6.1 below, we can show that, under conditions (6.6), the order of the variance of

$$\Delta_n(x) := \{L_n(x)\}^{-1} \sum_{i=1}^n \{\Phi(x\sqrt{1 + (\frac{X_i-\mu}{b_n})^2} - \frac{X_i-\mu}{b_n}) - \Phi(x)\}$$

is $\{n(1 - G(a_n) + G(-a_n))\}^{-2}$, which goes to ∞ rather than zero as $n \to \infty$. Hence, $\Delta_n(x)$ may not converge in probability to one. In order to estimate the error term $L_n(x)$, a more explicit expansion is needed. Based on (6.7) and (6.8), we derive the following exact limits under conditions (6.6). These limits show that the expansion for the Student's t-statistic depends on the tail quantities of F, which explains why $\Delta_n(x)$ does not converge in probability to one.

Theorem 6.1. *(Gong and Peng, 2010) Suppose conditions (6.6) hold. Then*

$$\lim_{n\to\infty} \frac{P(\sqrt{n}(\bar{X}_n-\mu)/\sigma)\le x)-\Phi(x)}{n\{1-G(a_n)+G(-a_n)\}} = l_3(x;p,\rho) \tag{6.9}$$

and

$$\lim_{n\to\infty} \frac{P(T_n\le x)-\Phi(x)}{n\{1-G(a_n)+G(-a_n)\}} = l_4(x;p,\rho) \tag{6.10}$$

uniformly in $x \in R$, where

$$\begin{aligned} & l_3(x;p,\rho) \\ = \; & \frac{1}{2\pi}\int_{-\infty}^{\infty} t^{-1}\{-sgn(t)|t|^{2-\rho}A_\rho\cos(tx)+|t|^{2-\rho}B_\rho\sin(tx) \\ & +t^2\frac{1-|t|^{-\rho}}{\rho}\sin(tx)\}e^{-t^2/2}dt-\frac{\Phi'(x)x}{\rho} \end{aligned}$$

and

$$\begin{aligned} l_4(x;p,\rho) = \; & \textstyle\int_0^\infty py^{\rho-2}\{\Phi'(x\sqrt{1+y^2}-y)-\Phi'(x)\}x(1+y^2)^{-1/2}ydy \\ & +\textstyle\int_0^\infty(1-p)y^{\rho-2}\{\Phi'(x\sqrt{1+y^2}+y)-\Phi'(x)\}x(1+y^2)^{-1/2}ydy \\ & -\textstyle\int_1^\infty py^{\rho-2}\{\Phi'(x\sqrt{1+y^2}-y)-\Phi'(x)\}dy \\ & +\textstyle\int_1^\infty(1-p)y^{\rho-2}\{\Phi'(x\sqrt{1+y^2}+y)-\Phi'(x)\}dy \\ & +\textstyle\int_1^\infty y^{\rho-2}\Phi'(x)x(1+y^2)^{-1/2}ydy \\ & +\textstyle\int_0^1 y^{\rho-2}\Phi'(x)x\{(1+y^2)^{-1/2}-1\}ydy \\ & -\textstyle\int_0^1 py^{\rho-2}\{\Phi'(x\sqrt{1+y^2}-y)-\Phi'(x)+\Phi''(x)y\}dy \\ & +\textstyle\int_0^1(1-p)y^{\rho-2}\{\Phi'(x\sqrt{1+y^2}+y)-\Phi'(x)-\Phi''(x)y\}dy \end{aligned}$$

Remark 6.1. *Under some refined conditions than (6.6), for example,*

$$1-G(x)+G(-x)=cx^{-2+\rho}\{1+O(x^{-\beta})\} \quad \text{and} \quad \frac{1-G(x)}{1-G(x)+G(-x)}=p+O(x^{-\beta})$$

for some $c>0$ and $\beta>0$, one can estimate $2-\rho$, $1-G(a_n)+G(-a_n)$ and p by a tail index estimator, a tail probability estimator and the estimator in de Haan and Pereira (1999), respectively. Hence we can estimate the leading error terms in (6.9) and (6.10). For more details on tail index estimation and tail probability estimation, we refer to de Haan and Ferreira (2006).

Next we give a detailed comparison study on these two limits. First, neither $l_3(x;p,\rho)$ nor $l_4(x;p,\rho)$ is an even function, which is different from the case when the third moment is finite. Secondly, we plot the difference $r(x;p,\rho)=\{l_3(x;p,\rho)\}^2-\{l_4(x;p,\rho)\}^2$ as a function of x and ρ for fixed $p=0.2,0.5,0.8$; see Figure 6.1. From the figure, we observe that i) $r(x;p,\rho)$ does not have a constant sign, which implies that one-sided confidence intervals based on $\sqrt{n}\{\bar{X}_n-\mu\}/\sigma$ are not always more accurate than those based on T_n; ii) when ρ is near zero, i.e., F has heavier tails, or $p=0.5$, the Student's t-statistic is preferred. Similar to Figures 6.1, we plot the difference

$$r^*(x;\rho)=\{l_3(x;p,\rho)-l_3(-x;p,\rho)\}^2-\{l_4(x;p,\rho)-l_4(-x;p,\rho)\}^2$$

as a function of $x>0$ and ρ in Figure 6.2. Note that it is easy to check that $r^*(x;\rho)$ defined above is independent of p. From Figure 6.2, we conclude that the Student's t-statistic is preferred for constructing two-sided confidence intervals.

6.3 Proofs

Proof of Theorem 6.1.

It is easy to check that

$$\begin{cases} \sigma^2=2\int_0^\infty\{1-G(x)+G(-x)\}x\,dx \\ 2n(a_n)^{-2}H(a_n)=1 \quad \text{for large} \quad n \\ \lim_{n\to\infty}2H(a_n)=\sigma^2. \end{cases} \tag{6.11}$$

Using the first and the second identity in (6.11), we have

$$\begin{aligned} \tfrac{\sigma\sqrt{n}}{a_n}-1 &= \{a_n^2(\tfrac{\sigma\sqrt{n}}{a_n}+1)\}^{-1}\{\sigma^2 n-a_n^2\} \\ &= \{\tfrac{\sigma\sqrt{n}}{a_n}+1\}^{-1}\{2n\textstyle\int_1^\infty(1-G(a_nx)+G(-a_nx))x\,dx\}. \end{aligned}$$

Hence, by (6.11) and Potter's bound (see Bingham, Goldie and Teugels (1987)), we have

$$\lim_{n\to\infty}\frac{\sigma\sqrt{n}/a_n-1}{n\{1-G(a_n)+G(-a_n)\}}=\int_1^\infty\lim_{n\to\infty}\frac{1-G(a_nx)+G(-a_n)}{1-G(a_n)+G(-a_n)}x\,dx=\frac{1}{-\rho}. \tag{6.12}$$

It follows from (6.7) and (6.12) that

$$\begin{aligned}
&\lim_{n\to\infty} \tfrac{P(\sqrt{n}(\bar{X}_n-\mu)/\sigma)\le x)-\Phi(x)}{n\{1-G(a_n)+G(-a_n)\}} \\
= &\lim_{n\to\infty} \tfrac{P(\sum_{i=1}^n (X_i-\mu)/a_n \le \frac{\sigma\sqrt{n}}{a_n}x)-\Phi(\frac{\sigma\sqrt{n}}{a_n}x)}{n\{1-G(a_n)+G(-a_n)\}} \\
&+\lim_{n\to\infty} \tfrac{\Phi(\frac{\sigma\sqrt{n}}{a_n}x)-\Phi(x)}{n\{1-G(a_n)+G(-a_n)\}} \\
= &\; l_3(x;p,\rho)
\end{aligned}$$

uniformly in $x \in R$.

Next, we show (6.10). It follows from (6.12) that

$$\sigma\sqrt{n}/a_n \to 1 \quad \text{as} \quad n\to\infty. \tag{6.13}$$

By (6.13) and (6.6),

$$\frac{n^{-1/2}}{n\{1-G(a_n)+G(-a_n)\}} \to 0 \quad \text{as} \quad n\to\infty. \tag{6.14}$$

Since $\lim_{n\to\infty} b_n^2/n = \sigma^2$, (6.13) implies that

$$a_n/b_n \to 1 \quad \text{as} \quad n\to\infty. \tag{6.15}$$

Write

$$\begin{aligned}
&b_n^{-1}E\{(X-\mu)I(|X-\mu|>b_n)\} \\
&= -\textstyle\int_1^\infty x d\{1-G(b_n x)\} + \int_1^\infty x dG(-b_n x) \\
&= 1-G(b_n)-G(-b_n)+\textstyle\int_1^\infty \{1-G(b_n x)-G(-b_n x)\}dx.
\end{aligned}$$

It follows from (6.6), (6.15) and Potter's bound that

$$\begin{aligned}
&\lim_{n\to\infty} \tfrac{nb_n^{-1}|E\{(X-\mu)I(|X-\mu|\le b_n)\}|}{n\{1-G(a_n)+G(-a_n)\}} \\
&= \lim_{n\to\infty} \tfrac{nb_n^{-1}|E\{(X-\mu)I(|X-\mu|> b_n)\}|}{n\{1-G(a_n)+G(-a_n)\}} \\
&= |2p-1|\tfrac{2-\rho}{1-\rho}.
\end{aligned}$$

Similarly, we have

$$\lim_{n\to\infty} \frac{nb_n^{-3}|E\{(X-\mu)^3 I(|X-\mu|\le b_n)\}|}{n\{1-G(a_n)+G(-a_n)\}} = |2p-1|\frac{2-\rho}{1+\rho}$$

and

$$\lim_{n\to\infty} \frac{nb_n^{-4}E\{(X-\mu)^4 I(|X-\mu|\le b_n)\}}{n\{1-G(a_n)+G(-a_n)\}} = \frac{2-\rho}{2+\rho}$$

Hence

$$\lim_{n\to\infty}\frac{\delta_n}{n\{1-G(a_n)+G(-a_n)\}}=1+|2p-1|\frac{2-\rho}{1-\rho}+|2p-1|\frac{2-\rho}{1+\rho}+\frac{2-\rho}{2+\rho} \quad (6.16)$$

Therefore, (6.14) and (6.16) imply that

$$\lim_{n\to\infty}\frac{P(T_n\le x)-\Phi(x)}{n\{1-G(a_n)+G(-a_n)\}}=\lim_{n\to\infty}\frac{L_n(x)}{n\{1-G(a_n)+G(-a_n)\}}.$$

Write

$$\begin{aligned}
L_n(x) &= n\textstyle\int_{-\infty}^{\infty}\{\Phi(x\sqrt{1+\frac{y^2}{b_n^2}}-\frac{y}{b_n})-\Phi(x)\}dG(y)\\
&= -n\textstyle\int_0^{\infty}\{\Phi(x\sqrt{1+\frac{y^2}{b_n^2}}-\frac{y}{b_n})-\Phi(x)\}d\{1-G(y)\}\\
&\quad -n\textstyle\int_0^{\infty}\{\Phi(x\sqrt{1+\frac{y^2}{b_n^2}}+\frac{y}{b_n})-\Phi(x)\}dG(-y)\\
&= n\textstyle\int_0^{\infty}\{1-G(y)\}\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}-\frac{y}{b_n})\{x(1+\frac{y^2}{b_n^2})^{-1/2}\frac{y}{b_n^2}-\frac{1}{b_n}\}dy\\
&\quad +n\textstyle\int_0^{\infty}G(-y)\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}+\frac{y}{b_n})\{x(1+\frac{y^2}{b_n^2})^{-1/2}\frac{y}{b_n^2}+\frac{1}{b_n}\}dy\\
&= n\textstyle\int_0^{\infty}\{1-G(y)\}\{\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}-\frac{y}{b_n})-\Phi'(x)\}x(1+\frac{y^2}{b_n^2})^{-1/2}\frac{y}{b_n^2}dy\\
&\quad +n\textstyle\int_0^{\infty}G(-y)\{\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}+\frac{y}{b_n})-\Phi'(x)\}x(1+\frac{y^2}{b_n^2})^{-1/2}\frac{y}{b_n^2}dy\\
&\quad -n\textstyle\int_0^{\infty}\{1-G(y)\}\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}-\frac{y}{b_n})\frac{1}{b_n}dy\\
&\quad +n\textstyle\int_0^{\infty}G(-y)\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}+\frac{y}{b_n})\frac{1}{b_n}dy\\
&\quad +n\textstyle\int_0^{\infty}\{1-G(y)\}\Phi'(x)x(1+\frac{y^2}{b_n^2})^{-1/2}\frac{y}{b_n^2}dy\\
&\quad +n\textstyle\int_0^{\infty}G(-y)\Phi'(x)x(1+\frac{y^2}{b_n^2})^{-1/2}\frac{y}{b_n^2}dy
\end{aligned}$$

$$\begin{aligned}
&= n\textstyle\int_0^{\infty}\{1-G(y)\}\{\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}-\frac{y}{b_n})-\Phi'(x)\}x(1+\frac{y^2}{b_n^2})^{-1/2}\frac{y}{b_n^2}dy\\
&\quad +n\textstyle\int_0^{\infty}G(-y)\{\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}+\frac{y}{b_n})-\Phi'(x)\}x(1+\frac{y^2}{b_n^2})^{-1/2}\frac{y}{b_n^2}dy\\
&\quad -n\textstyle\int_{b_n}^{\infty}\{1-G(y)\}\{\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}-\frac{y}{b_n})-\Phi'(x)\}\frac{1}{b_n}dy\\
&\quad +n\textstyle\int_{b_n}^{\infty}G(-y)\{\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}+\frac{y}{b_n})-\Phi'(x)\}\frac{1}{b_n}dy\\
&\quad -n\textstyle\int_0^{\infty}\{1-G(y)-G(-y)\}\Phi'(x)\frac{1}{b_n}dy\\
&\quad +n\textstyle\int_{b_n}^{\infty}\{1-G(y)+G(-y)\}\Phi'(x)x(1+\frac{y^2}{b_n^2})^{-1/2}\frac{y}{b_n^2}dy\\
&\quad -n\textstyle\int_0^{b_n}\{1-G(y)\}\{\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}-\frac{y}{b_n})-\Phi'(x)-\Phi'(x)x\frac{y}{b_n}\}\frac{1}{b_n}dy\\
&\quad +n\textstyle\int_0^{b_n}G(-y)\{\Phi'(x\sqrt{1+\frac{y^2}{b_n^2}}+\frac{y}{b_n})-\Phi'(x)+\Phi'(x)x\frac{y}{b_n}\}\frac{1}{b_n}dy\\
&\quad +n\textstyle\int_0^{b_n}\{1-G(y)+G(-y)\}\Phi'(x)x\{(1+\frac{y^2}{b_n^2})^{-1/2}-1\}\frac{y}{b_n^2}dy\\
&= I_1+\cdots+I_9.
\end{aligned}$$

Using (6.6), (6.15) and Potter's bound, it is easy to check that

$$
\begin{aligned}
&\tfrac{I_1}{n\{1-G(a_n)+G(-a_n)\}} \\
&= \int_0^\infty \tfrac{1-G(b_n y)}{1-G(a_n)+G(-a_n)} \{\Phi'(x\sqrt{1+y^2}-y) - \Phi'(x)\} x(1+y^2)^{-1/2} y dy \\
&\to \int_0^\infty p y^{\rho-2} \{\Phi'(x\sqrt{1+y^2}-y) - \Phi'(x)\} x(1+y^2)^{-1/2} y dy.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\tfrac{I_2}{n\{1-G(a_n)+G(-a_n)\}} &\to \int_0^\infty (1-p) y^{\rho-2} \{\Phi'(x\sqrt{1+y^2}+y) - \Phi'(x)\} x(1+y^2)^{-1/2} y dy, \\
\tfrac{I_3}{n\{1-G(a_n)+G(-a_n)\}} &\to -\int_1^\infty p y^{\rho-2} \{\Phi'(x\sqrt{1+y^2}-y) - \Phi'(x)\} dy, \\
\tfrac{I_4}{n\{1-G(a_n)+G(-a_n)\}} &\to \int_1^\infty (1-p) y^{\rho-2} \{\Phi'(x\sqrt{1+y^2}+y) - \Phi'(x)\} dy, \\
\tfrac{I_6}{n\{1-G(a_n)+G(-a_n)\}} &\to \int_1^\infty y^{\rho-2} \Phi'(x) x(1+y^2)^{-1/2} y dy, \\
\tfrac{I_7}{n\{1-G(a_n)+G(-a_n)\}} &\to -\int_0^1 p y^{\rho-2} \{\Phi'(x\sqrt{1+y^2}-y) - \Phi'(x) + \Phi''(x) y\} dy, \\
\tfrac{I_8}{n\{1-G(a_n)+G(-a_n)\}} &\to \int_0^1 (1-p) y^{\rho-2} \{\Phi'(x\sqrt{1+y^2}+y) - \Phi'(x) - \Phi''(x) y\} dy, \\
\tfrac{I_9}{n\{1-G(a_n)+G(-a_n)\}} &\to \int_0^1 y^{\rho-2} \Phi'(x) x\{(1+y^2)^{-1/2} - 1\} y dy,
\end{aligned}
$$

and

$$
I_5 \quad = -n\Phi'(x) \frac{1}{b_n} E(X-\mu) = 0.
$$

Note that, for I_7 and I_8, we use the fact that $\Phi'(x)x = -\Phi''(x)$. Hence, (6.10) follows from the above equations. □

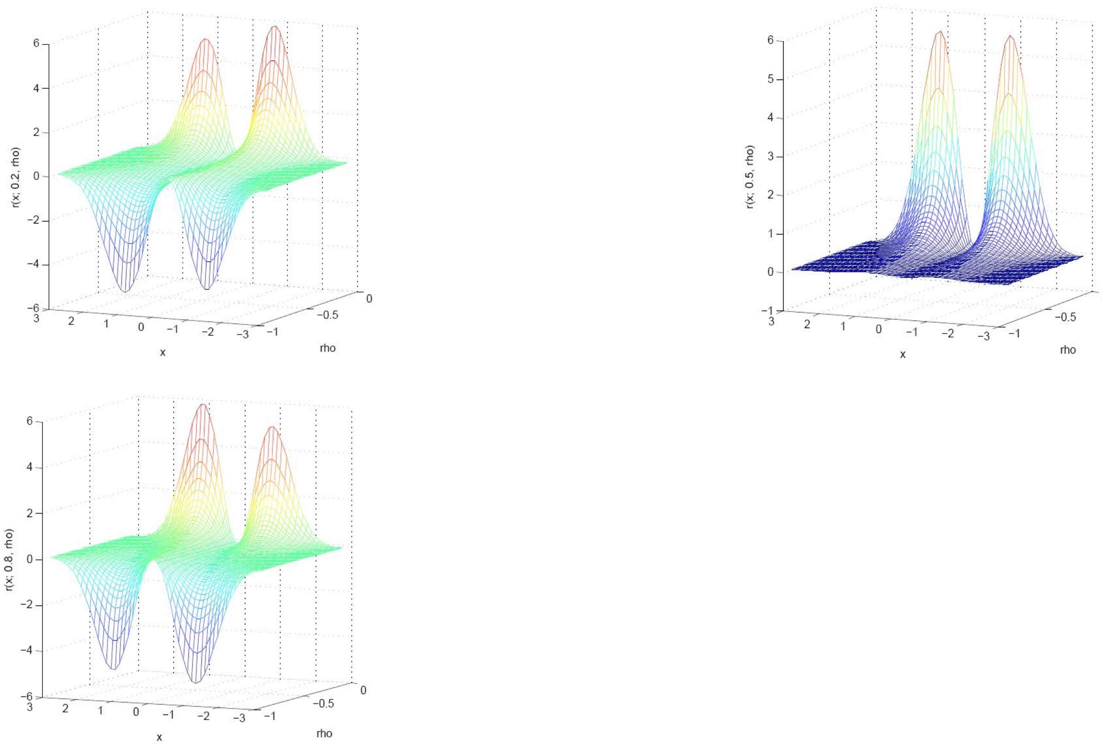

Figure 6.1: Plots of $r(x; p, \rho)$ as a function of x and ρ.

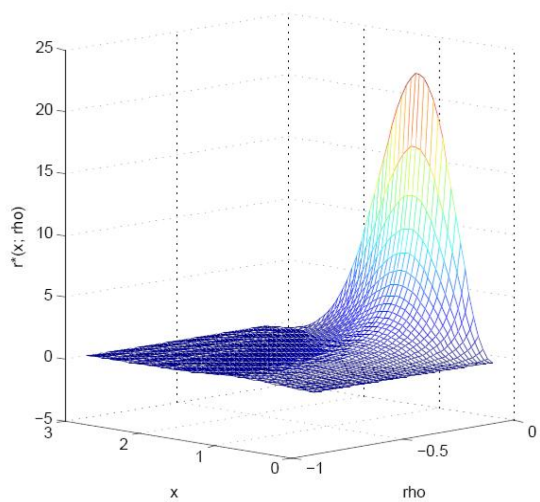

Figure 6.2: Plot of $r^*(x; \rho)$ as a function of x and ρ.

CHAPTER VII

A SPECIFICATION TEST FOR NESTED STOCHASTIC MODELS WITH DISCRETE AND DEPENDENT OBSERVATIONS

7.1 Introduction

Stochastic processes give a natural framework to model financial quantities such as the returns of traded assets, or some hidden factors in a financial model such as the market price of risk, instantaneous volatility in a stochastic volatility model, etc. Both discrete-time processes and continuous-time processes have been extensively studied in the finance literature. A good reference for discrete-time models is Hamilton (1994). See also Teräsvirta (2009) for reviews on more recent univariate models of conditional heteroskedasticity. Shreve (2004) is a good introduction to continuous-time models.

Aït-Sahalia (1996a) proposed a parametric specification test to examine whether a particular model should be accepted by comparing the model-implied density with the empirical one which is estimated nonparametrically. In practice, a common focus in modeling is on whether a particular subset of parameters is significant. For example, if the observed short-term interest rate time series can be explained by a jump-diffusion model with regime shifting, a natural question to ask could be whether the jump part or the regime shifting part of the process is significant. Normally, the maximum likelihood (MLE) ratio test is used to test the significance of the parameters. However, there are cases where MLE could be difficult to carry out, for example due to the difficulty or cost of computing the transition density. Li (2010) also pointed out through numerical simulation that with finite sample size, there are cases where MLE could fail to detect the damped diffusion function while fabricates nonlinear drift function. Li (2010) uses an alternative method based on Aït-Sahalia's parametric specification test which is able to detect the damped diffusion function.

In this chapter, we give a more systematic and rigorous study on the method proposed in Li (2010). Similar to the log-likelihood ratio statistic, the proposed statistic is the difference of two statistics in the two nested models, where each test statistic measures the distance of the theoretical marginal density with the one implied by the model and estimated using a nonparametric kernel estimation. For more information on kernel estimation technique, see for example, Silverman (1986) and Wand and Jones (1994). A general weight function is introduced in the test statistic to allow for flexibility and more control in detecting fine structures in the data. Under similar regularity conditions as in Aït-Sahalia (1996a), we derive the asymptotic distribution of the proposed statistic. A major tool used in our derivation is the convergence properties of the empirical process. Dependency in the data plays a very crucial role in the asymptotic distribution here.

We then carry out a simulation study for the Vasicek model for which the theoretical asymptotic limit can be analytically computed. Through parametric bootstrapping, we compare the sizes and powers of the proposed parametric specification test with maximum likelihood ratio test. The results show that in this special multi-normality case, these two tests have very similar sizes and powers. This simulation also establishes bootstrapping as a reliable and accurate method of computing the critical value of the test statistic given a specific significance level.

Next we apply the proposed parametric specification test on Chicago Board Options Exchange's daily VIX volatility index. Two different nested models are studied: the asymmetric piece-wise linear stochastic elasticity variance model (ALSEV), and Aït-Sahalia's (1996a) non-linear stochastic elasticity variance model (NLSEV). In both models, the restricted model is that the drift function is exactly linear. Interestingly, while maximum likelihood ratio tests cannot reject the linearity of the drift function, the proposed test rejects linearity under both models. This study shows that the proposed test could be very useful, especially in situations where maximum likelihood might be susceptible to finite sample bias. The content of this chapter is based on Y. Gong, M. Li and L. Peng (2011), A specification test for nested stochastic models with discrete and dependent observations.

7.2 Parametric specification test for nested models

7.2.1 Null hypothesis and assumptions

Let $\{X_t, t \geq 0\}$ be an observed strictly stationary data sequence with marginal density $\pi(x;\theta)$ where the parameter vector $\theta = (\theta_1, \cdots, \theta_d)^T \in \Omega \subset R^d$ and Ω is compact, and $x \in D = (\underline{x}, \bar{x})$, with $-\infty \leq \underline{x} < \bar{x} \leq \infty$.

Let $\Omega_0 = \{(\theta_1, \cdots, \theta_d)^T \in \Omega : \theta_1 = \theta_{1,0}, \cdots, \theta_m = \theta_{m,0}\}$, where $1 \leq m \leq d$ and $\theta_{1,0}, \cdots, \theta_{m,0}$ are given. Assume $\theta_0 \in \Omega$ is the true parameter.

We are interested in testing the following hypothesis:

$$H_0 : \theta \in \Omega_0 \ \text{ vs. } \ H_a : \theta \in \Omega \backslash \Omega_0.$$

Some regularity conditions are stated in the following assumptions.

Assumption 7.1. *In a neighborhood of θ_0, $\pi(x, \theta)$ is twice continuously differentiable in θ, $E[(\partial\pi(x,\theta)/\partial\theta)(\partial\pi(x,\theta)/\partial\theta^T)]$ has full rank, and $\partial^2\pi(x,\theta)/\partial\theta_i\partial\theta_j$ and $\partial^3\pi(x,\theta)/\partial\theta_i\partial\theta_j\partial\theta_k$ are bounded in absolute value for all $\theta \in \Omega$, $x \in D$, i and j. For true parameter θ_0, $u(x)$, $u'(x)$, $\partial^3\pi(x,\theta_0)/\partial\theta_i\partial\theta_j\partial\theta_k$, $\frac{\partial}{\partial\theta}\pi(x;\theta_0)$ and $\frac{\partial}{\partial x}\{u(x)\frac{\partial}{\partial\theta}\pi(x;\theta_0)\}$ are integrable, where $u(x)$ is defined in (7.7). $u(x)$, $\pi(x;\theta_0)$, $\frac{\partial^2}{\partial x^2}\pi(x;\theta_0)$ and $\frac{\partial}{\partial x}\pi(x;\theta_0)$ are bounded on $x \in D$.*

Assumption 7.2. *Starting from any point in the interior of D, the boundaries $\underline{x}$ and $\bar{x}$ cannot be attained in finite expected time.*

Assumption 7.3. *$\{X_t, t \geq 0\}$ is a β-mixing sequence with mixing bound $\beta_k = O(\lambda^k)$, for some $0 < \lambda < 1$.*

One important example of our observed data sequence $\{X_t, t \geq 0\}$ is the diffusion process determined by Itô stochastic differential equations:

$$dX_t = \mu(X_t, \theta)dt + \sigma(X_t, \theta)dW_t, \tag{7.1}$$

where $\theta = (\theta_1, \cdots, \theta_d)^T \in \Omega \subset R^d$ and Ω is compact, and $\{W_t, t \geq 0\}$ is a standard Brownian motion. The functions $\mu(\cdot, \theta)$ and $\sigma^2(\cdot, \theta)$ are the drift and the diffusion functions of the process. In this particular case, the marginal density has the following explicit expression

$$\pi(x; \theta) = \frac{\xi(\theta)}{\sigma^2(x, \theta)} \exp\left\{\int_{x_0}^{x} \frac{2\mu(u, \theta)}{\sigma^2(u, \theta)} du\right\}, \tag{7.2}$$

where x_0 is an arbitrary interior point in D and $\xi(\theta)$ is a normalization constant depending on x_0 so that $\pi(x; \theta)$ integrates to 1.

In order to guarantee that (7.1) admits a unique strong solution with a stationary marginal density as well as Assumption 7.2 and Assumption 7.3, we can make the following assumptions

Assumption 7.4. *For every $\theta \in \Omega$:*

1. *$\mu(x, \theta)$ and $\sigma(x, \theta)$ are six times continuously differentiable in x.*

2. *The integral of $m(v, \theta) \equiv (1/\sigma^2(v, \theta)) \exp\{-\int_v^{\epsilon_0} 2\mu(u, \theta)/\sigma^2(u, \theta) du\}$ converges at both boundaries of D.*

3. *The integral of $s(v, \theta) \equiv \exp\{\int_v^{\epsilon_0} 2\mu(u, \theta)/\sigma^2(u, \theta) du\}$ diverges at both boundaries of D.*

where ϵ_0 is any fixed point in D.

Assumption 7.5. $\lim_{x \to \underline{x} \text{ or } x \to \bar{x}} \sigma(x, \theta)\pi(x, \theta) = 0$ *and* $\lim_{x \to \underline{x} \text{ or } x \to \bar{x}} |\sigma(x, \theta)/\{2\mu(x, \theta) - \sigma(x, \theta)\frac{\partial}{\partial x}\sigma(x, \theta)\}| < \infty$.

Assumption 7.5 guarantees the following mixing bound for the observations: $2\alpha_k \leq \beta_k \leq \lambda^k/2$, for some $0 < \lambda < 1$. See Aït-Sahalia (1996b).

Assumption 7.6. *Let K be a continuously twice differentiable symmetric density with support in $[-1, 1]$. Let $h = h_n > 0$ be a bandwidth chosen such that, for some $0 < \iota < 1$, $nh^{3+\iota} \log^{-4} n \to \infty$ and $nh^4 \to 0$, as $n \to \infty$.*

7.2.2 Properties of the empirical process

Let $F(x) = P(X \leq x)$ denote the marginal distribution function of X with marginal density $\pi(x; \theta)$. Given discretely observed dependent data $X = \{X_i\}$ with $i = 1, 2, \cdots, n$, let $F_n(x)$ be the empirical distribution function with n observations. That is,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i \leq x}. \tag{7.3}$$

The empirical process has the following well-known convergence property:

Lemma 7.1. *Define $\beta_n(x) \equiv \sqrt{n}\{F_n(x) - F(x)\}$ to be the empirical process of the distribution function F. Under Assumption 7.3, we have*

$$\beta_n(x) \xrightarrow{D} W(x),$$

where $W(\cdot)$ is a Gaussian process with zero mean and covariance structure

$$\phi(x, y) \equiv EW(x)W(y) = \phi_0(x, y) + \phi_a(x, y) \tag{7.4}$$

with

$$\phi_0(x, y) = F(x \wedge y) - F(x)F(y); \tag{7.5}$$

$$\phi_a(x, y) = \sum_{j=1}^{\infty} \phi_j(x, y) + \sum_{j=1}^{\infty} \phi_j(y, x), \tag{7.6}$$

and where $\phi_j(x, y) = \mathrm{Cov}\{I(X_1 \leq x), I(X_{j+1} \leq y)\}$.

Results similar to the above are often generally called central limit theorems. An i.i.d. version of the above lemma was first proved by Donsker (1952), confirming an early conjecture by Doob that the limit is a Brownian bridge process. The result has subsequently been relaxed to various types of weakly dependent data, see for example, Billingsley (1968). For a theoretical introduction to the empirical process technique, see for example, Kosorok (2008).

With assumptions above, the empirical process has the following additional property, which we need later.

Lemma 7.2. *(Zhou 1996) Under Assumption 7.2 and 7.3, for any $0 < \iota < 1$, we have*

$$\sup_{|s-t|\leq h} |\beta_n(t) - \beta_n(s)| = O(h^{(1-\iota)/2} \log^2 n) \quad a.s.,$$

where h satisfies the condition in Assumption 7.6.

7.2.3 Test statistic and its asymptotic distribution

Define

$$L(\theta; X) = \sum_{i=1}^{n} u(X_i)\{\pi(X_i;\theta) - \hat{\pi}(X_i)\}^2, \tag{7.7}$$

where $u(\cdot) > 0$ is a bounded and continuously differentiable weight function and

$$\hat{\pi}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{7.8}$$

is the estimator of the marginal density. Here the kernel K and bandwidth h are assumed to satisfy Assumption 7.6 above. Define

$$\tilde{\theta} = \arg\min_{\theta\in\Omega_0} L(\theta; X), \quad \hat{\theta} = \arg\min_{\theta\in\Omega} L(\theta; X). \tag{7.9}$$

Lemma 7.3. *(Aït-Sahalia 1996a) Under H_0, $\tilde{\theta}$ and $\hat{\theta}$ are, respectively, consistent and asymptotically normal estimators of the true parameter θ_0, i.e.,*

$$\tilde{\theta} \xrightarrow{p} \theta_0 \quad and \quad \hat{\theta} \xrightarrow{p} \theta_0,$$

and

$$\|\tilde{\theta} - \theta_0\| = \|\hat{\theta} - \theta_0\| = O_p(n^{-1/2}).$$

Our test statistic is defined as

$$\hat{M} = L(\tilde{\theta}; X) - L(\hat{\theta}; X). \tag{7.10}$$

The following theorem gives the asymptotic distribution of $\hat{M}$.

Theorem 7.1. *(Gong, Li and Peng 2011) Under H_0, the test statistic $\hat{M}$ is distributed asymptotically as*

$$\hat{M} \xrightarrow{d} \frac{1}{2} N^T (A^{-1} - H) N, \tag{7.11}$$

where

$$A = 2\,\mathbb{E}\left[u(X)\frac{\partial}{\partial\theta}\pi(X;\theta_0)\frac{\partial}{\partial\theta^T}\pi(X;\theta_0)\right] \equiv \begin{pmatrix} A_1 & A_2 \\ A_2^T & A_3 \end{pmatrix}, \tag{7.12}$$

with the expectation $\mathbb{E}$ taken under the stationary density measure $\pi(\cdot\ ;\theta)$, and

$$H = \begin{pmatrix} 0_{m\times m} & 0_{m\times(d-m)} \\ 0_{(d-m)\times m} & A_3^{-1} \end{pmatrix}.$$

The vector N is a Gaussian random vector with $EN = 0$ and covariance structure

$$V_N = \int\int \phi(x,y)\frac{\partial^2}{\partial x\partial\theta}\{u(x)\pi^2(x;\theta_0)\}\frac{\partial^2}{\partial y\partial\theta^T}\{u(y)\pi^2(y;\theta_0)\}dxdy. \tag{7.13}$$

Remark 7.1. *In the special case of independently observed data, $\phi_a(x,y) \equiv 0$, and the covariance matrix V_N simplifies to*

$$V_N = \mathrm{Cov}\left(\frac{\partial}{\partial\theta}\{u(X)\pi^2(X;\theta_0)\}\right). \tag{7.14}$$

In general, in addition to $\phi_0(x,y)$, the autocorrelation term $\phi_a(x,y)$ also contributes to V_N which is when Theorem 7.1 becomes more interesting and useful.

It's useful to look at the above test statistic in the special case of one restriction equation with weight function $u(\cdot) \equiv 1$ for independently observed data and with diagonal matrix A. The following corollary shows that in this case, the proposed test statistic has a familiar $\chi^2(1)$ asymptotic distribution, up to a constant. This allows us to quickly compute the critical values or confidence intervals. The proof of the corollary follows the above remark easily and is omitted.

Corollary 7.1. *Let the discretely observed data be independently generated from the stationary density $\pi(x;\theta)$ which has a diagonal matrix A. Let $u(\cdot) \equiv 1$. Let the null hypothesis be $H_0 : \theta_1 = \theta_{1,0}$. Under Assumptions 7.1 to 7.6, the test statistic $\hat{M}$ is asymptotically distributed as*

$$\hat{M} \xrightarrow{d} C \cdot \chi^2(1), \tag{7.15}$$

where the constant C is given by

$$C = \frac{1}{4}\left[\mathbb{E}\Big(\frac{\partial}{\partial\theta_1}\pi(X;\theta_0)\Big)^2\right]^{-1} \cdot \mathrm{Var}\left[\frac{\partial}{\partial\theta_1}\pi^2(X;\theta_0)\right]. \tag{7.16}$$

In the general case of dependent observations and with non-diagonal matrix A, we need to perform the integral in Theorem 7.1 in order to obtain the asymptotic distribution. This in practice could either be difficult or costly. Bootstrapping gives very convenient alternative to compute, for example, the critical value. In the next section, we establish the accuracy of bootstrapping method through the special case of Vasicek model where the discrete observations are multivariate normally distributed.

7.3 Simulation study

We now examine the finite sample performance of the test statistic proposed in the last section. The approach we take is to compare the theoretical asymptotic distribution with the empirical distribution from parametric bootstrapping. In order to have the asymptotic distribution in closed form, we consider the simple Vasicek model (1977) in the class of one-dimensional diffusion processes:

$$dX_t = \kappa(m - X_t)dt + \sigma dW_t, \tag{7.17}$$

where $\kappa > 0$, $m > 0$ and $\sigma > 0$ are the parameters and W_t is a Brownian motion. Notice that for all diffusion processes, the shape of the stationary density does not uniquely determine the drift and diffusion functions. For example, if $\mu(X_t)$ and $\sigma(X_t)$ are the drift and diffusion functions, then a new diffusion process with drift function $4\mu(X_t)$ and diffusion function $2\sigma(X_t)$ would give exactly the same stationary density, as can be seen from equation (7.2). This indeterminateness is also discussed in Li (2010). Thus we treat σ as a nuisance parameter which is to be estimated using some other methods. Let $\theta = (m, \kappa)$ in the full model. The restricted model is the Vasicek model with fixed long-run mean. The simple null hypothesis we want to test is then $H_0 : m = m_0$ for some fixed $m_0 > 0$. This example is very simple-minded, but it does allow us to have a closed-form asymptotic distribution in order to examine the finite sample performance of the proposed test statistic.

The Vasicek model has a closed-form solution as follows:

$$X_t = X_s e^{-\kappa(t-s)} + m(1 - e^{-\kappa(t-s)}) + \sigma \int_s^t e^{-\kappa(t-u)} dW_u, \tag{7.18}$$

for any $t > s$. In the steady state, X is normally distributed with mean m and variance $\Sigma \equiv \sigma^2/2\kappa$. Therefore, the cumulative distribution for the invariant measure is given by

$$F(x) = \Phi\Big(\frac{x - m}{\sqrt{\Sigma}}\Big). \tag{7.19}$$

Assuming that observations are made a constant δ apart. Then, for $j \geq 0$,

$$\phi_j(x, y) = \phi_j(y, x) = \Phi\left(\frac{x - m}{\Sigma}, \frac{y - m}{\Sigma}, e^{-j\delta\kappa}\right) - \Phi\left(\frac{x - m}{\Sigma}\right)\Phi\left(\frac{y - m}{\Sigma}\right), \tag{7.20}$$

where $\Phi(\cdot, \cdot, \rho)$ is the bivariate cumulative normal distribution function with correlation coefficient ρ, and $\Phi(\cdot)$ is the univariate cumulative normal distribution function.

To compute $\phi_a(x, y)$, we need to sum over $\phi_j(x, y)$ in equation (7.20), which is very slowly converging. In the section 7.5, we show that

$$\phi_a(x,y) = \varphi\left(\frac{x-m}{\Sigma}\right)\varphi\left(\frac{y-m}{\Sigma}\right)\sum_{\ell=1}^{\infty}\frac{2}{\ell!}\frac{e^{-\delta\kappa\ell}}{1-e^{-\delta\kappa\ell}}H_{\ell-1}\left(\frac{x-m}{\Sigma}\right)H_{\ell-1}\left(\frac{y-m}{\Sigma}\right), \tag{7.21}$$

where $\varphi(\cdot)$ is the standard normal density function and $H_\ell(\cdot)$ is the one-dimensional Hermite polynomial with order ℓ. This alternative expression converges extremely fast, as is shown in Figure 7.1. As we see, the method using Hermite polynomials converges very fast with few summation terms, while the method using bivariate cumulative normal distribution requires hundreds of terms in order to achieve relatively high accuracy.

It's useful to take a look at the functions $\phi_0(x,y)$ and $\phi(x,y)$ in Theorem 7.1. Figure 7.2 shows the surface and contour plots of these two functions for the Vasicek model. The underlying parameters are: $\kappa = 0.1685$, $m = 0.0582$, and $\sigma = 0.0186$. These plots show that autocorrelation terms are very important in $\phi(x,y)$ as the peak of $\phi_0(x,y)$ is only about one tenth of that of $\phi(x,y)$. Also, as the contour plots show, the dependency structures in $\phi(x,y)$ and $\phi_0(x,y)$ are quite different.

Specializing Theorem 7.1 to the Vasicek model, and utilizing the tractability of the Vasicek model, we have the following theorem:

Theorem 7.2. *Consider the Vasicek model above with discrete data observed a constant δ time apart. Under $H_0 : m = m_0$ for some fixed $m_0 > 0$, the test statistic $\hat{M}$ is asymptotically distributed as*

$$\hat{M} \xrightarrow{d} C \cdot \chi^2(1), \tag{7.22}$$

where the constant C is given by

$$C = \frac{3\sqrt{3}}{5\sqrt{5}\pi}\frac{\kappa}{\sigma^2} + \frac{3\sqrt{3}}{4\pi}\frac{\kappa}{\sigma^2}\sum_{j=1}^{\infty}\frac{e^{-\kappa\delta(2j-1)}}{1-e^{-\kappa\delta(2j-1)}}\frac{\left[(2j-1)!!\right]^2}{(2j-1)!}\left(\frac{2}{3}\right)^{2j+1}. \tag{7.23}$$

Remark 7.2. *The method used to derive the above Theorem is not restricted to Vasicek model, but rather applicable to any Gaussian diffusion processes (such as the* *log-Ornstein-Uhlenbeck process), or any discrete time series models where observations are jointly normally distributed.*

The benchmark parameters we use are $\kappa = 0.1685$, $m = 0.0582$, and $\sigma = 0.0186$ taken from Li (2010). These are obtained from the maximum likelihood estimation for the monthly federal funds rate from July 1954 to June 2008. There are 648 total observations. For a visualization of the data, we refer readers to Li (2010). Three different data generating processes are used with the above κ and σ values but different values of m. More specifically, we let $m = 0.0482$, $m = 0.0582$ and $m = 0.0682$. The null hypothesis we test is $H_0 : m = 0.0582$ for all three data generating processes. When $m = 0.0582$, we are interested in whether the proposed test gives correct empirical size. When $m = 0.0482$ and $m = 0.0682$, we are interested in the empirical power of the proposed test to rejected the incorrect null hypothesis.

Table 7.1 reports the means of the parameters estimated from the proposed parametric specification test (PST) in this paper as well as the maximum likelihood estimation (MLE) for two nested Vasicek models. For MLE, the theoretical transition density is available in closed form for Vasicek model. For PST, following Aït-Sahalia (1996a), we use a Gaussian Kernel when estimating the empirical marginal density with the bandwidth also chosen in a similar way to Aït-Sahalia (1996a). Three different data generating processes are used with the same κ and σ values but different values of m. More specifically, we let $m = 0.0482$, $m = 0.0582$ and $m = 0.0682$ in three different cases but fix $\kappa = 0.1685$ and $\sigma = 0.0186$ to generate 1000 sample paths for each case. We use Euler scheme to generated time series with a frequency of 12 points per day, but the final data points are sampled with a time interval of one month and a total length of 648. Using other schemes such as Milstein scheme gives little difference in the simulated time series because of the small time intervals and the dimensionality of 1. For each simulated path, in the full model, both m and κ are estimated while in the restricted model m is fixed at $m = 0.0582$. The standard deviations of the 1000 estimated parameters are in parentheses. As we can see that for full model two methods give almost the same estimates for the parameter m, and PST method provides better estimate for the parameter κ.

Table 7.2 reports the empirical powers and sizes of the proposed parametric specification test (PST) in this paper as well as the maximum likelihood estimation (MLE) for two nested Vasicek models. Three different data generating processes are used with the same κ and σ values but different values of m. More specifically, we let $m = 0.0482$, $m = 0.0582$ and $m = 0.0682$ in three different cases but fix $\kappa = 0.1685$ and $\sigma = 0.0186$ to generate 1000 sample paths for each case. The null hypothesis is $H_0 : m = 0.0582$. For each sample path, the proposed test statistic and the maximum likelihood ratio are computed. The empirical powers or sizes are computed as the proportions of test statistics exceeding the theoretical critical values (from Theorem 7.2) for the corresponding quantile ($\alpha = 5\%$ or 10%). As we can see from the table, MLE method has slightly larger power than PST, whereas PST has more accurate size than MLE.

While Theorem 7.1 provides the theoretical asymptotic distribution for the proposed test statistic $\hat{M}$, in practice it is often convenient to use a bootstrapping method to compute the critical values. Figure 7.3 shows the histograms from bootstrapping and Theorem 1. The top subplot gives the histogram of the theoretical asymptotic distribution (constant multiple of a chi-squared distribution of degree 1) with 100,000 draws. The bottom subplot is the histogram of the test statistic estimated from 1000 simulations of the Vasicek model. The bootstrapping method gives an empirical distribution that's fairly close to the theoretical distribution, and the 5% critical values from the two methods are very close.

7.4 Empirical tests on CBOE's VIX index

7.4.1 Data

The data we use in the empirical test is the daily CBOE VIX index from January 2, 1990 to November 9, 2009.

Figure 7.4 plots the data. The top subplot graphs the VIX index levels (normalized to percentage by dividing it by 100). There are a total of 5004 daily observations. A noticeable feature is the high levels during the subprime mortgage crisis in years 2007-2009. The bottom subplot graphs the daily VIX changes.

7.4.2 Asymmetric piece-wise linear stochastic elasticity variance model (ALSEV)

The model we are interested in is specified by the following diffusion process:

$$dX_t = \Big[\kappa(m - X_t) + \delta\big|m - X_t\big|\Big]dt + \sqrt{\beta_1 X_t + \beta_2 X_t^{\beta_3}}\, dW_t. \tag{7.24}$$

Here, $\theta = (\kappa, m, \delta, \beta_1, \beta_2, \beta_3)$ is the parameter vector, and W_t is a standard Brownian motion. Here we assume that $\kappa \pm \delta > 0$, $m > 0$, and $\beta_1 X_t + \beta_2 X_t^{\beta_3} \geq 0$ on the domain of X_t so that the process is well-defined and well-behaved.

The null hypothesis is $H_0 : \delta = 0$, that is, there is no asymmetry for the drift strength below or above the level m. The existence of nonlinearity for the drift function in diffusion processes has generated considerable interest. For example, Aït-Sahalia (1996a) proposes a flexible specification for the short-rate data in which the drift is a nonlinear function of the state variable. Also, by using various estimation methods, Chapman and Pearson (2000), Jones (2003), and Li et al. (2004) all prevent evidence that the nonlinearity in the drift function might be spurious for the short-rate data. See also Ang and Bekaert (2002), Durham (2003), and Takamizawa (2008). Through extensive Monte Carlo simulation, Li (2010) presents evidence that for finite sample size, maximum likelihood estimation often cannot distinguish the effect of a nonlinear drift function and a damped diffusion function. For volatility index data, both nonlinear and linear drift diffusion processes have been studied, see for example, Bakshi, Ju, and Ou-Yang (2005), and Dotsis, Psychoyios, and Skiadopoulos (2007).

The specification for the drift function in equation (7.24) is in different form from Aït-Sahalia (1996a), although both specifications relax the assumption of global linearity. There are two attractive features of this proposed form. First, it adds only one additional parameter as opposed to two in Aït-Sahalia (1996a). Second, it is piece-wise linear and globally Lipschitz. Therefore, it is a natural nesting model to consider when studying whether a globally linear drift specification is sufficient. The slight disadvantage of this specification is that its derivative has a discontinuity in the middle.

Table 7.3 reports estimation results from the parametric specification test (PST) proposed in this paper (with weight function $u(\cdot) \equiv 1$) as well as the maximum likelihood estimation (MLE) for the asymmetric piece-wise linear stochastic elasticity model (ALSEV) as in equation (7.24). The diffusion parameters are fixed at $\beta_1 = -0.0371$, $\beta_2 = 1.998$ and $\beta_3 = 2.397$. The proposed test statistic as well as the maximum likelihood ratio are reported. The transition probabilities needed in MLE is computed through PDE method. The critical value c_α (with $\alpha = 5\%$) is computed from a bootstrapping method similar to the one used the simulation test. In contrast to the maximum likelihood estimation, the parametric specification test rejects the null hypothesis of a globally linear drift.

7.4.3 Aït-Sahalia's non-linear stochastic elasticity variance model (NL-SEV)

The model we study next is identical to the one proposed in Aït-Sahalia (1996a), but written in a slight different form:

$$dX_t = \left[\kappa(m - X_t) + \delta_1 \left(\frac{1}{X_t} - \frac{1}{m}\right) + \delta_2(m^2 - X_t^2)\right] dt + \sqrt{\beta_1 X_t + \beta_2 X_t^{\beta_3}}\, dW_t. \tag{7.25}$$

Here $\theta = (\kappa, m, \delta_1, \delta_2, \beta_1, \beta_2, \beta_3)$ is the parameter vector with $\delta_1 > 0$ and $\delta_2 > 0$ to guarantee mean reversion when X_t is close to 0 or very large. Notice that there is no requirement for the positivity for κ since the linear term will be dominated by one of the nonlinear terms in either case. Compared with the original form $\mu(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_{-1}/x$, the current form makes the position of the nonlinear mean reversions m more explicit. To transform to the original form, notice

$$\alpha_0 = \kappa m - \delta_1/m + \delta_2 m^2, \tag{7.26}$$

$$\alpha_1 = -\kappa, \tag{7.27}$$

$$\alpha_2 = -\delta_2, \tag{7.28}$$

$$\alpha_{-1} = \delta_1. \tag{7.29}$$

The null hypothesis is $H_0 : \delta_1 = \delta_2 = 0$, that is, there is no nonlinear mean reversion below or above the level m.

Table 7.4 reports estimation results from the parametric specification test (PST) proposed in this paper (with weight function $u(\cdot) \equiv 1$) as well as the maximum likelihood estimation (MLE) for the asymmetric piece-wise linear stochastic elasticity model (ALSEV) as in equation (7.24). The diffusion parameters are treated as nuisance parameters and fixed at $\beta_1 = -0.0371$, $\beta_2 = 1.998$ and $\beta_3 = 2.397$. The proposed test statistic as well as the maximum likelihood ratio are reported. The transition probabilities needed in MLE is computed through PDE method. The critical value c_α (with $\alpha = 5\%$) is computed from a bootstrapping method similar to the one used in the simulation test. In contrast to the maximum likelihood estimation, the parametric specification test rejects the null hypothesis of a globally linear drift.

7.5 Proofs

Proof of Theorem 7.1. By Taylor expansion,

$$\pi(x;\theta) = \pi(x;\theta_0) + \frac{\partial}{\partial\theta^T}\pi(x;\theta_0)(\theta-\theta_0) + O(\|\theta-\theta_0\|^2), \tag{7.30}$$

where $O(\cdot)$ holds uniformly for $x \in D$. Also, we have the following facts,

$$\begin{aligned} & \sup_{\theta:\|\theta-\theta_0\|=O(n^{-1/2})} \|\tfrac{1}{n}\textstyle\sum_{i=1}^n \frac{\partial}{\partial\theta}\pi(X_i;\theta) - \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta}\pi(X_i;\theta_0)\| \\ = \quad & O_p(\|\theta-\theta_0\|) = O_p(n^{-1/2}), \end{aligned} \tag{7.31}$$

which follows from the mean value theorem, and

$$\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta}\pi(X_i;\theta_0) \xrightarrow{p} E\frac{\partial}{\partial\theta}\pi(x;\theta_0). \tag{7.32}$$

Using Lemma 7.2, it can be shown that,

$$\frac{1}{n}\sum_{i=1}^n \{\hat\pi(X_i) - \pi(X_i,\theta_0)\}u(X_i) = \int\{\hat\pi(s) - \pi(s,\theta_0)\}u(s)dF_n(s) = o_p(1), \tag{7.33}$$

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n \{\hat\pi(X_i) - \pi(X_i,\theta_0)\}u(X_i)\frac{\partial^2}{\partial\theta\partial\theta^T}\pi(X_i;\theta_0) \\ & \qquad = \int\{\hat\pi(s) - \pi(s,\theta_0)\}u(s)\frac{\partial^2}{\partial\theta\partial\theta^T}\pi(s;\theta_0)dF_n(s) = o_p(1), \end{aligned} \tag{7.34}$$

which followed by the uniform consistency of kernel estimator.

By Lemma 7.3, (7.33) and (7.30), we can write

$$\begin{aligned} \hat M &= L(\tilde\theta;X) - L(\hat\theta;X) \\ &= \textstyle\sum_{i=1}^n u(X_i)\{\pi(X_i;\tilde\theta) - \pi(X_i;\hat\theta)\}\{\pi(X_i;\tilde\theta) - \hat\pi(X_i) + \pi(X_i;\hat\theta) - \hat\pi(X_i)\} \\ &= \textstyle\sum_{i=1}^n u(X_i)\{\frac{\partial}{\partial\theta^T}\pi(X_i;\theta_0)(\tilde\theta-\hat\theta) + O_p(n^{-1})\}\{2\{\pi(X_i;\theta_0) - \hat\pi(X_i)\} \\ &\quad + \tfrac{\partial}{\partial\theta^T}\pi(X_i;\theta_0)(\tilde\theta-\theta_0) + \tfrac{\partial}{\partial\theta^T}\pi(X_i;\theta_0)(\hat\theta-\theta_0) + O_p(n^{-1})\} \\ &= 2\textstyle\sum_{i=1}^n u(X_i)\{\frac{\partial}{\partial\theta^T}\pi(X_i;\theta_0)(\tilde\theta-\hat\theta)\}\{\pi(X_i;\theta_0) - \hat\pi(X_i)\} \\ &\quad + \textstyle\sum_{i=1}^n u(X_i)(\tilde\theta-\theta_0)^T\frac{\partial}{\partial\theta}\pi(X_i;\theta_0)\frac{\partial}{\partial\theta^T}\pi(X_i;\theta_0)(\tilde\theta-\theta_0) \\ &\quad - \textstyle\sum_{i=1}^n u(X_i)(\hat\theta-\theta_0)^T\frac{\partial}{\partial\theta}\pi(X_i;\theta_0)\frac{\partial}{\partial\theta^T}\pi(X_i;\theta_0)(\hat\theta-\theta_0) + o_p(1) \\ &= 2\textstyle\sum_{i=1}^n u(X_i)\{\frac{\partial}{\partial\theta^T}\pi(X_i;\theta_0)(\tilde\theta-\hat\theta)\}\{\pi(X_i;\theta_0) - \hat\pi(X_i)\} + n(\tilde\theta-\theta_0)^T D_n(\tilde\theta-\theta_0) \\ &\quad - n(\hat\theta-\theta_0)^T D_n(\hat\theta-\theta_0) + o_p(1) \\ &= \tfrac{\partial}{\partial\theta^T}L(\theta_0;X)(\tilde\theta-\hat\theta) + n(\tilde\theta-\theta_0)^T D_n(\tilde\theta-\theta_0) - n(\hat\theta-\theta_0)^T D_n(\hat\theta-\theta_0) + o_p(1) \end{aligned} \tag{7.35}$$

where $D_n = \frac{1}{n}\sum_{i=1}^n u(X_i)\frac{\partial}{\partial\theta}\pi(X_i;\theta_0)\frac{\partial}{\partial\theta^T}\pi(X_i;\theta_0)$, and

$$D_n \xrightarrow{p} Eu(x)\frac{\partial}{\partial\theta}\pi(x;\theta_0)\frac{\partial}{\partial\theta^T}\pi(x;\theta_0) = \frac{1}{2}A. \tag{7.36}$$

Next, we are going to find an expression for $\tilde\theta-\hat\theta$, $\tilde\theta-\theta_0$ and $\hat\theta-\theta_0$ in (7.35).

First, we need the following facts. By definition, we have

$$\begin{aligned}
&\tfrac{1}{n}\tfrac{\partial^2}{\partial\theta\partial\theta^T}L(\theta;X)\\
=\ &\tfrac{2}{n}\textstyle\sum_{i=1}^n u(X_i)\tfrac{\partial}{\partial\theta}\pi(X_i;\theta)\tfrac{\partial}{\partial\theta^T}\pi(X_i;\theta)+\tfrac{2}{n}\textstyle\sum_{i=1}^n u(X_i)\{\pi(X;\theta)-\hat\pi(X_i)\}\tfrac{\partial^2\pi(X_i;\theta)}{\partial\theta\partial\theta^T}.
\end{aligned} \tag{7.37}$$

Hence, using (7.31), (7.32) and (7.34), it's easy to check that

$$\sup_{\|\theta-\theta_0\|=o(1)}\|\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta^T}L(\theta;X)-\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta^T}L(\theta_0;X)\|=o_p(1) \tag{7.38}$$

and

$$\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta^T}L(\theta_0;X)\xrightarrow{p}A. \tag{7.39}$$

Since, by the definition of $\hat\theta$, (7.38) and (7.39),

$$\begin{aligned}
\tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\tilde\theta;X) &= \tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\hat\theta;X)+\tfrac{1}{n}\tfrac{\partial^2}{\partial\theta\partial\theta^T}L(\theta^*;X)(\tilde\theta-\hat\theta)\\
&= \tfrac{1}{n}\tfrac{\partial^2}{\partial\theta\partial\theta^T}L(\theta^*;X)(\tilde\theta-\hat\theta)\\
&= A(\tilde\theta-\hat\theta)+o_p(\|\tilde\theta-\hat\theta\|),
\end{aligned} \tag{7.40}$$

where θ^* lies between $\tilde\theta$ and $\hat\theta$, we have

$$\tilde\theta-\hat\theta=A^{-1}\{\frac{1}{n}\frac{\partial}{\partial\theta}L(\tilde\theta;X)\}+o_p(\|\tilde\theta-\hat\theta\|). \tag{7.41}$$

Again by (7.38) and (7.39), we can write

$$\begin{aligned}
\tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\tilde\theta;X) &= \tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\theta_0;X)+\tfrac{1}{n}\tfrac{\partial^2}{\partial\theta\partial\theta^T}L(\theta^*;X)(\tilde\theta-\theta_0)\\
&= \tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\theta_0;X)+A(\tilde\theta-\theta_0)+o_p(\|\tilde\theta-\theta_0\|),
\end{aligned} \tag{7.42}$$

where θ^* lies between $\tilde\theta$ and θ_0.

Since the last $d-m$ elements in $\frac{\partial}{\partial\theta}L(\tilde\theta;X)$ are zero and the first m elements in $\tilde\theta-\theta_0$ are zero under H_0, (7.42) implies that

$$\begin{aligned}
0 &= H\tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\tilde\theta;X)\\
&= H\tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\theta_0;X)+HA(\tilde\theta-\theta_0)+o_p(\|\tilde\theta-\theta_0\|)\\
&= H\tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\theta_0;X)+(\tilde\theta-\theta_0)+o_p(\|\tilde\theta-\theta_0\|),
\end{aligned}$$

i.e.,

$$\tilde{\theta} - \theta_0 = -H\frac{1}{n}\frac{\partial}{\partial\theta}L(\theta_0; X) + o_p(\|\tilde{\theta} - \theta_0\|). \tag{7.43}$$

By (7.42) and (7.43), we have

$$\frac{1}{n}\frac{\partial}{\partial\theta}L(\tilde{\theta}; X) = (I_{d\times d} - AH)\frac{1}{n}\frac{\partial}{\partial\theta}L(\theta_0; X) + o_p(\|\tilde{\theta} - \theta_0\|). \tag{7.44}$$

Hence, by (7.41) and (7.44), we have

$$\tilde{\theta} - \hat{\theta} = (A^{-1} - H)\frac{1}{n}\frac{\partial}{\partial\theta}L(\theta_0; X) + o_p(\|\tilde{\theta} - \theta_0\|) + o_p(\|\tilde{\theta} - \hat{\theta}\|). \tag{7.45}$$

Also, from

$$0 = \frac{1}{n}\frac{\partial}{\partial\theta}L(\hat{\theta}; X) = \frac{1}{n}\frac{\partial}{\partial\theta}L(\theta_0; X) + A(\hat{\theta} - \theta_0) + o_p(\|\hat{\theta} - \theta_0\|),$$

we have

$$\hat{\theta} - \theta_0 = -A^{-1}\frac{1}{n}\frac{\partial}{\partial\theta}L(\theta_0; X) + o_p(\|\hat{\theta} - \theta_0\|). \tag{7.46}$$

Hence, by (7.43), (7.45) and (7.46), we have from (7.35),

$$\begin{aligned}
\hat{M} &= \tfrac{1}{n}\{\tfrac{\partial}{\partial\theta}L(\theta_0; X)\}^T(A^{-1} - H)\{\tfrac{\partial}{\partial\theta}L(\theta_0; X)\} + n\{\tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\theta_0; X)\}^T H(\tfrac{A}{2})H\{\tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\theta_0; X)\} \\
&\quad - n\{\tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\theta_0; X)\}^T A^{-1}(\tfrac{A}{2})A^{-1}\{\tfrac{1}{n}\tfrac{\partial}{\partial\theta}L(\theta_0; X)\} + o_p(1) \\
&= \tfrac{1}{2}\{\tfrac{1}{\sqrt{n}}\tfrac{\partial}{\partial\theta}L(\theta_0; X)\}^T(A^{-1} - 2H + HAH)\{\tfrac{1}{\sqrt{n}}\tfrac{\partial}{\partial\theta}L(\theta_0; X)\} + o_p(1) \\
&= \tfrac{1}{2}\{\tfrac{1}{\sqrt{n}}\tfrac{\partial}{\partial\theta}L(\theta_0; X)\}^T(A^{-1} - H)\{\tfrac{1}{\sqrt{n}}\tfrac{\partial}{\partial\theta}L(\theta_0; X)\} + o_p(1).
\end{aligned} \tag{7.47}$$

Next, we are going to find the limit distribution for $\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}L(\theta_0; X)$, which finally gives the limit distribution for $\hat{M}$.

Write

$$\hat{\pi}(s) = \frac{1}{h}\int K(\frac{s-t}{h})\, dF_n(t).$$

We then have

$$\begin{aligned}
\hat{\pi}(s) - \pi(s; \theta_0) &= \textstyle\int \frac{1}{h}K(\frac{s-t}{h})dF_n(t) - \pi(s; \theta_0) \\
&= \textstyle\int \frac{1}{h}K(\frac{s-t}{h})d\{F_n(t) - F(t)\} + \int K(t)\{\pi(s - ht; \theta_0) - \pi(s; \theta_0)\}dt \\
&:= I_1(s) + I_2(s).
\end{aligned} \tag{7.48}$$

Also, write

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}L(\theta_0;X) = 2\int \sqrt{n}\{\pi(s;\theta_0) - \hat{\pi}(s)\}u(s)\frac{\partial}{\partial\theta}\pi(s;\theta_0)dF_n(s).$$

Then, using (7.48), we get

$$\begin{aligned}
\tfrac{1}{\sqrt{n}}\tfrac{\partial}{\partial\theta}L(\theta_0;X) &= 2\textstyle\int \sqrt{n}\{\pi(s;\theta_0) - \hat{\pi}(s)\}u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0)dF_n(s) \\
&= -2\textstyle\int \sqrt{n}\{I_1(s) + I_2(s)\}u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0)dF(s) \\
&\quad - 2\textstyle\int \sqrt{n}\{I_1(s) + I_2(s)\}u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0)d\{F_n(s) - F(s)\} \\
&= -2\textstyle\int \sqrt{n}I_1(s)u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0)dF(s) - 2\textstyle\int \sqrt{n}I_2(s)u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0)dF(s) \\
&\quad - 2\textstyle\int \sqrt{n}I_1(s)u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0)d\{F_n(s) - F(s)\} \\
&\quad - 2\textstyle\int \sqrt{n}I_2(s)u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0)d\{F_n(s) - F(s)\} \\
&:= -2B_1 - 2B_2 - 2B_3 - 2B_4.
\end{aligned} \tag{7.49}$$

By Lemmas 7.1, 7.2, and integration by parts, we have

$$\begin{aligned}
B_1 &= \textstyle\int\int \frac{1}{h}K(\frac{s-t}{h})u(s)\pi(s;\theta_0)\frac{\partial}{\partial\theta}\pi(s;\theta_0)dsd\{\sqrt{n}(F_n(t) - F(t))\} \\
&= \textstyle\int\int K(s)u(t+sh)\pi(t+sh;\theta_0)\frac{\partial}{\partial\theta}\pi(t+sh;\theta_0)dsd\{\sqrt{n}(F_n(t) - F(t))\} \\
&= -\textstyle\int\int K(s)\frac{\partial}{\partial t}\{u(t+sh)\pi(t+sh;\theta_0)\frac{\partial}{\partial\theta}\pi(t+sh;\theta_0)\}\sqrt{n}\{F_n(t) - F(t)\}dsdt \\
&= -\textstyle\int\int K(s)\frac{\partial}{\partial t}\{u(t)\pi(t;\theta_0)\frac{\partial}{\partial\theta}\pi(t;\theta_0)\} \\
&\quad \times \{\{\beta_n(t-sh) - \beta_n(t)\} + \beta_n(t)\}dsdt \\
&\xrightarrow{d} -\textstyle\int W(t)\frac{\partial}{\partial t}\{u(t)\pi(t;\theta_0)\frac{\partial}{\partial\theta}\pi(t;\theta_0)\}dt,
\end{aligned}$$

where W is a Gaussian process with zero mean and covariance structure $EW(x)W(y) = \phi(x,y)$. In the above, we require $\frac{\partial}{\partial t}\{u(t)\pi(t;\theta_0)\frac{\partial}{\partial\theta}\pi(t;\theta_0)\}$ to be integrable.

Similarly,

$$\begin{aligned}
B_2 &= \textstyle\int \sqrt{n}\int K(t)\{\pi(s-ht;\theta_0) - \pi(s;\theta_0)\}dtu(s)\frac{\partial}{\partial\theta}\pi(s;\theta_0)dF(s) \\
&= \textstyle\int \sqrt{n}\int K(t)\{\frac{\partial}{\partial s}\pi(s;\theta_0)(-ht) + \frac{1}{2}\frac{\partial^2}{\partial s^2}\pi(s^*;\theta_0)h^2t^2\}dtu(s)\frac{\partial}{\partial\theta}\pi(s;\theta_0)dF(s) \\
&= \tfrac{\sqrt{n}h^2}{2}\textstyle\int\int K(t)\frac{\partial^2}{\partial s^2}\pi(s^*;\theta_0)t^2\}dtu(s)\frac{\partial}{\partial\theta}\pi(s;\theta_0)dF(s) \\
&\to 0.
\end{aligned}$$

In the above, we require $\frac{\partial^2}{\partial s^2}\pi(s;\theta_0)$ to be bounded on $s \in D$ and $nh^4 \to 0$.

After using integration by parts twice, we obtain

$$
\begin{aligned}
B_3 &= -\int\int \tfrac{\sqrt{n}}{h^2}\{F_n(t)-F(t)\}\{F_n(s)-F(s)\}\{\tfrac{1}{h}K''(\tfrac{s-t}{h})u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0) \\
&\quad + K'(\tfrac{s-t}{h})\tfrac{\partial}{\partial s}\{u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0)\}\}dsdt \\
&= -\int\int \tfrac{\sqrt{n}}{h^2}\{F_n(t)-F(t)\}\{F_n(t+hz)-F(t+hz)\}\{K''(z)u(t+hz)\tfrac{\partial}{\partial\theta}\pi(t+hz;\theta_0) \\
&\quad + hK'(z)\tfrac{\partial}{\partial t}\{u(t+hz)\tfrac{\partial}{\partial\theta}\pi(t+hz;\theta_0)\}\}dzdt \\
&:= -\int\int \tfrac{\sqrt{n}}{h^2}\{F_n(t)-F(t)\}\{F_n(t+hz)-F(t+hz)\}b(z,t,h)dzdt \\
&= -\int\int \tfrac{1}{\sqrt{n}h^2}\{\beta_n(t)\{\beta_n(t+hz)-\beta_n(t)\}+\beta_n^2(t)\}b(z,t,h)dzdt \\
&= O_p(\tfrac{h^{(1-\iota)/2}\log^2 n}{\sqrt{n}h^2}) + O_p(\tfrac{h}{\sqrt{n}h^2}) \\
&\to 0.
\end{aligned}
$$

Note that the third term $O_p(\frac{h}{\sqrt{n}h^2})$ is due to the fact that $\int\int b(z,t,h)dzdt = O_p(h)$ which follows from Taylor expansion. In the above, we require $\frac{\partial}{\partial\theta}\pi(x;\theta_0)$ and $\frac{\partial}{\partial x}\{u(x)\frac{\partial}{\partial\theta}\pi(x;\theta_0)\}$ to be integrable and bounded. It's easy to check that

$$
\begin{aligned}
B_4 &= -\int K(t)\beta_n(t)\{\{\tfrac{\partial}{\partial s}\pi(s-ht;\theta_0) - \tfrac{\partial}{\partial s}\pi(s;\theta_0)\}u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0) \\
&\quad + \{\pi(s-ht;\theta_0)-\pi(s;\theta_0)\}\tfrac{\partial}{\partial s}\{u(s)\tfrac{\partial}{\partial\theta}\pi(s;\theta_0)\}\}dsdt \\
&= O_p(h) \to 0
\end{aligned}
$$

In the above, we require $\frac{\partial^2}{\partial s^2}\pi(s;\theta_0)$ and $\frac{\partial}{\partial s}\pi(s;\theta_0)$ to be bounded, and $u(s)\frac{\partial}{\partial\theta}\pi(s;\theta_0)$ to be integrable.

Therefore, from (7.49), we have

$$
\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}L(\theta_0;X) \xrightarrow{d} 2\int W(t)\frac{\partial}{\partial t}\{u(t)\pi(t;\theta_0)\frac{\partial}{\partial\theta}\pi(t;\theta_0)\}dt := N.
$$

Hence, from (7.47), we get

$$
\hat{M} \xrightarrow{d} \frac{1}{2}N^T(A^{-1}-H)N,
$$

where N is a Gaussian random vector with $EN=0$ and covariance structure

$$
\begin{aligned}
V_N \equiv ENN^T &= 4\int\int \phi(t,s)\tfrac{\partial}{\partial t}\{u(t)\pi(t;\theta_0)\tfrac{\partial}{\partial\theta}\pi(t;\theta_0)\}\tfrac{\partial}{\partial s}\{u(s)\pi(s;\theta_0)\tfrac{\partial}{\partial\theta^T}\pi(s;\theta_0)\}dtds \\
&= \int\int \phi(t,s)\tfrac{\partial^2}{\partial t\partial\theta}\{u(t)\pi^2(t;\theta_0)\}\tfrac{\partial^2}{\partial s\partial\theta^T}\{u(s)\pi^2(s;\theta_0)\}dtds.
\end{aligned}
$$

This finished the proof for Theorem 7.1. $\square$

Proof of Remark following Theorem 7.1. To show the statement for independently observed data, it suffices to use repeated integration by parts. For example,

$$\begin{aligned}
&\iint F(x \wedge y) \frac{\partial^2}{\partial x \partial \theta}\{u(x)\pi^2(x;\theta_0)\} \frac{\partial^2}{\partial y \partial \theta^T}\{u(y)\pi^2(y;\theta_0)\} dx dy \\
&= -\iint \frac{\partial^2}{\partial x \partial \theta}\{u(x)\pi^2(x;\theta_0)\} dx \left(\frac{\partial}{\partial \theta^T}\{u(y)\pi^2(y;\theta_0)\}\right) \; d_y F(x \wedge y) \\
&= -\int \int_y^{\bar{x}} \frac{\partial^2}{\partial x \partial \theta}\{u(x)\pi^2(x;\theta_0)\} dx \; \pi(y;\theta_0) \left(\frac{\partial}{\partial \theta^T}\{u(y)\pi^2(y;\theta_0)\}\right) dy \\
&= \int \left(\frac{\partial}{\partial \theta}\{u(y)\pi^2(y;\theta_0)\}\right) \pi(y;\theta_0) \left(\frac{\partial}{\partial \theta^T}\{u(y)\pi^2(y;\theta_0)\}\right) dy \\
&= \mathbb{E} \frac{\partial \{u(X)\pi^2(X;\theta)\}}{\partial \theta} \cdot \frac{\partial \{u(X)\pi^2(X;\theta)\}}{\partial \theta^T}. \qquad (7.50)
\end{aligned}$$

$\square$

Proof of Theorem 7.2. From the explicit expression for the stationary density

$$\pi(x;\theta_0) = \sqrt{\frac{\kappa}{\pi \sigma^2}} \; e^{-\frac{\kappa(m-x)^2}{\sigma^2}} \qquad (7.51)$$

for the Vasicek model, we can easily compute

$$c \equiv (A^{-1} - H)_{11} = \frac{3\sqrt{3}\pi}{4} \frac{\sigma^4}{\kappa^2}. \qquad (7.52)$$

It remains to compute V_N. Let $V_N = V_0 + V_a$, where V_0 and V_a are the contributions of V_N from $\phi_0(x,y)$ and $\phi_a(x,y)$, respectively.

It is easy to compute V_0. By symmetry,

$$\mathbb{E}\left(\frac{\partial \pi^2(x;\theta_0)}{\partial m}\right) = 0, \qquad (7.53)$$

therefore

$$V_0 = \mathrm{Var}\left(\frac{\partial \pi^2(x;\theta_0)}{\partial m}\right) = \mathbb{E}\left(\frac{\partial \pi^2(x;\theta_0)}{\partial m}\right)^2 = \frac{8\kappa^3}{5\sqrt{5}\pi^2\sigma^6}. \qquad (7.54)$$

The term V_a is given by

$$V_a = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_a(x,y)\frac{\partial^2}{\partial x \partial m}\{\pi^2(x;\theta_0)\}\frac{\partial^2}{\partial y \partial m}\{\pi^2(y;\theta_0)\}dxdy. \tag{7.55}$$

For any $j \neq 0$, by the joint normality of discrete observations from Vasicek model as can be seen from equation (7.18), we have

$$\phi_j(x,y) = \Phi\left(\frac{x-m}{\Sigma}, \frac{y-m}{\Sigma}, e^{-j\delta\kappa}\right) - \Phi\left(\frac{x-m}{\Sigma}\right)\Phi\left(\frac{y-m}{\Sigma}\right), \tag{7.56}$$

where $\Phi(\cdot,\cdot,\rho)$ is the bivariate cumulative normal distribution function with correlation coefficient ρ, and $\Phi(\cdot)$ is the univariate cumulative normal distribution function.

We could perform the integral V_a by using the summation of cumulative bivariate-normal distributions. However, ϕ_a is very slowly converging in j. In order to perform the integral in V_j more efficiently, we perform a series expansion of $\phi_j(x,y)$ in the correlation coefficient $\rho = e^{-j\delta\kappa}$ to get

$$\phi_j(x,y) = \varphi\left(\frac{x-m}{\Sigma}\right)\varphi\left(\frac{y-m}{\Sigma}\right)\sum_{\ell=1}^{\infty}\frac{e^{-j\delta\kappa\ell}}{\ell!}H_{\ell-1}\left(\frac{x-m}{\Sigma}\right)H_{\ell-1}\left(\frac{y-m}{\Sigma}\right), \tag{7.57}$$

where $\varphi(\cdot)$ is the standard normal density function and $H_\ell(\cdot)$ is the one-dimensional Hermite polynomial with order ℓ. Therefore,

$$\phi_a(x,y) = 2\sum_{j=1}^{\infty}\phi_j(x,y) \tag{7.58}$$

$$= \varphi\left(\frac{x-m}{\Sigma}\right)\varphi\left(\frac{y-m}{\Sigma}\right)\sum_{\ell=1}^{\infty}\frac{2}{\ell!}\frac{e^{-\delta\kappa\ell}}{1-e^{-\delta\kappa\ell}}H_{\ell-1}\left(\frac{x-m}{\Sigma}\right)H_{\ell-1}\left(\frac{y-m}{\Sigma}\right). \tag{7.59}$$

Notice that this expansion has a very nice property. Although x and y cannot be separated in $\phi_a(x,y)$, in the above equation, x and y are separated in each of the terms in the summation. Therefore, the double integral in V_a reduces to a series of one-dimensional integrations

$$V_a = \sum_{\ell=1}^{\infty}\frac{2}{\ell!}\frac{e^{-\delta\kappa\ell}}{1-e^{-\delta\kappa\ell}}\,\mathbf{I}_\ell^2, \tag{7.60}$$

with

$$\mathbf{I}_\ell \equiv \int_{-\infty}^{\infty} \varphi\left(\frac{x-m}{\Sigma}\right) H_{\ell-1}\left(\frac{x-m}{\Sigma}\right) \frac{\partial^2}{\partial x \partial m}\{\pi^2(x;\theta_0)\}\, dx. \tag{7.61}$$

A simple computation gives that $\mathbf{I}_\ell = 0$ if ℓ is even, and if $\ell = 2j-1$ is odd,

$$\mathbf{I}_\ell = (-1)^{j+1}\left(\frac{2}{3}\right)^{j+1/2} \frac{(2j-1)!!}{\pi} \frac{\sqrt{\kappa}}{\sigma^3}. \tag{7.62}$$

Therefore,

$$V_a = \sum_{j=1}^{\infty} \frac{2}{(2j-1)!} \frac{e^{-\delta\kappa(2j-1)}}{1-e^{-\delta\kappa(2j-1)}} \left(\frac{2}{3}\right)^{2j+1} \frac{[(2j-1)!!]^2}{\pi^2} \frac{\kappa}{\sigma^6}. \tag{7.63}$$

This finishes the proof of Theorem 7.2 by noticing that $C = c(V_0 + V_a)/2$. $\square$

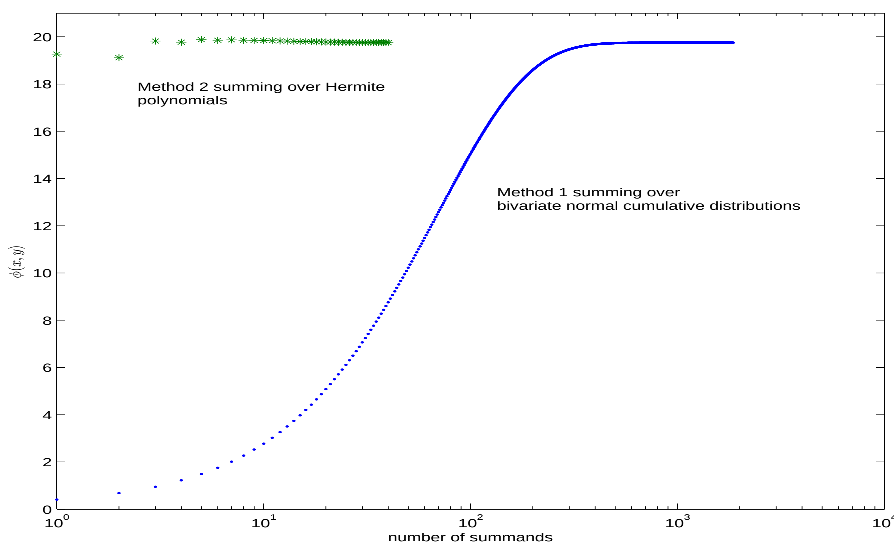

Figure 7.1: Comparison of speed of convergence for the two methods of computing $\phi(x, y)$ in the Vasicek model. The method using Hermite polynomials converges very fast with few summation terms, while the method using bivariate cumulative normal distribution requires hundreds of terms in order to achieve relatively high accuracy.

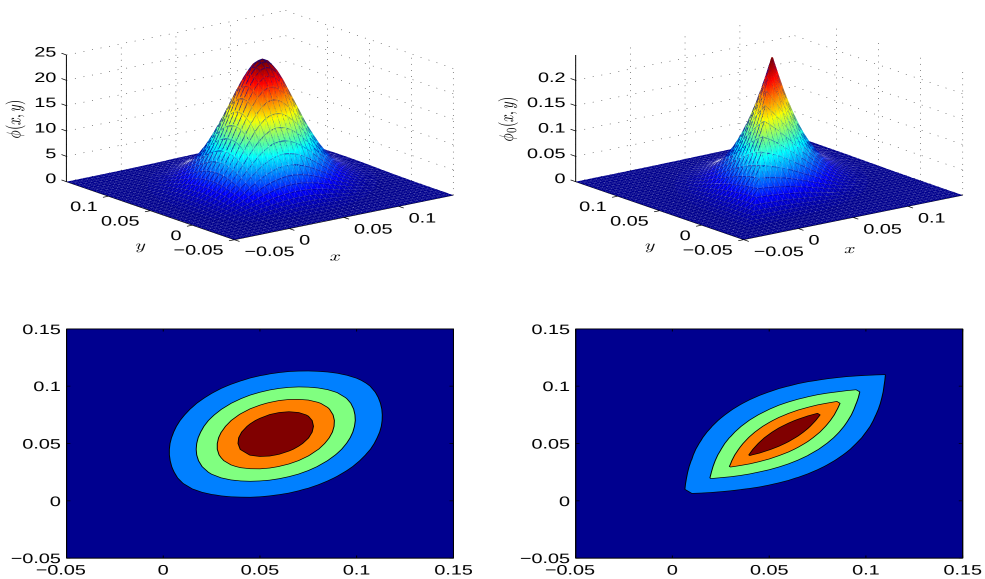

Figure 7.2: The functions $\phi(x, y)$ and $\phi_0(x, y)$ in Theorem 1 for Vasicek model. The left two subplots are the surface plot (top subplot) and contour plot (bottom subplot) of $\phi(x, y)$, while the right two subplots are those of $\phi_0(x, y)$. The underlying parameters are: $\kappa = 0.1685$, $m = 0.0582$, and $\sigma = 0.0186$. These plots show that autocorrelation terms are very important in $\phi(x, y)$ as the peak of $\phi_0(x, y)$ is only about one tenth of that of $\phi(x, y)$. Also, as the contour plots show, the dependency structures in $\phi(x, y)$ and $\phi_0(x, y)$ are quite different.

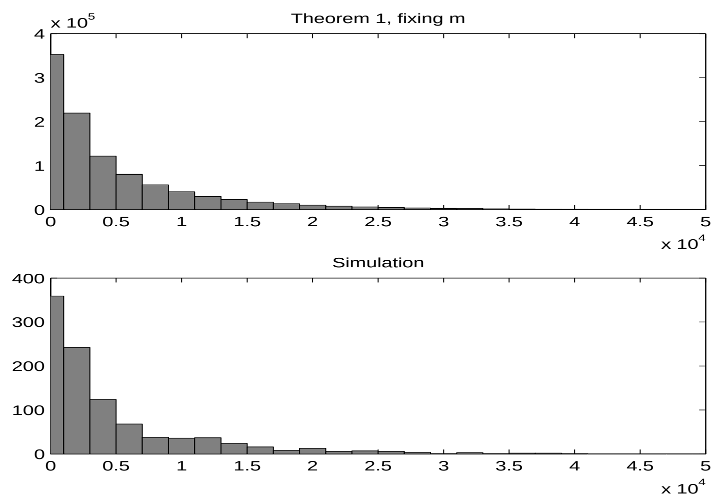

Figure 7.3: Histograms of test statistic from Theorem 1 and from bootstrapping method. The simulations use the Vasicek model with $\kappa = 0.1685$, $m = 0.0582$, and $\sigma = 0.0186$. The top subplot gives the histogram of the theoretical asymptotic distribution (constant multiple of a chi-squared distribution of degree 1) with 100,000 draws. The bottom subplot is the histogram of the test statistic estimated from 1000 simulations of the Vasicek model.

Table 7.1: Parameter estimates from the parametric specification test and maximum likelihood estimation for nested Vasicek models

		True data generating long-run mean					
		$m = 0.0482$		$m = 0.0582$		$m = 0.0682$	
		m	κ	m	κ	m	κ
Full	PST	0.0482	0.1734	0.0589	0.1726	0.0682	0.1716
		(0.009)	(0.044)	(0.009)	(0.044)	(0.010)	(0.044)
	MLE	0.0481	0.1927	0.0589	0.1906	0.0682	0.1918
		(0.009)	(0.049)	(0.008)	(0.049)	(0.009)	(0.049)
Restricted	PST		0.1509		0.1615		0.1485
			(0.042)		(0.042)		(0.042)
	MLE		0.1655		0.1786		0.1637
			(0.047)		(0.045)		(0.046)

Table 7.2: Empirical powers and sizes of the parametric specification test and maximum likelihood estimation for nested Vasicek models

		True data generating long-run mean		
$H_0: m = 0.0582$		$m = 0.0482$	$m = 0.0582$	$m = 0.0682$
	model	power	size	power
$\alpha = 5\%$	PST	16.0%	5.1%	17.9%
	MLE	16.6%	6.2%	18.7%
$\alpha = 10\%$	PST	29.6%	9.6%	29.7%
	MLE	30.3%	11.0%	30.8%

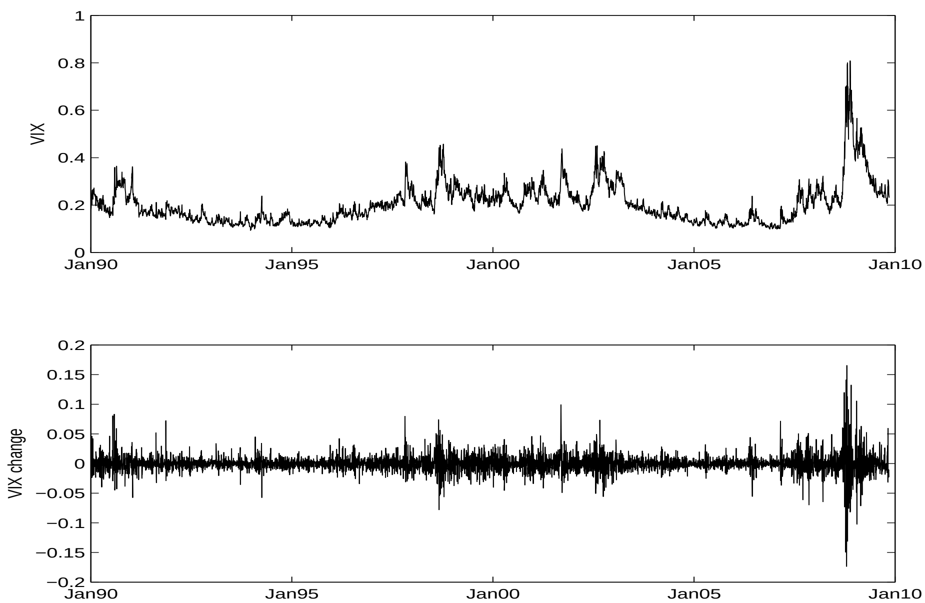

Figure 7.4: Daily CBOE VIX index levels and changes from January 2, 1990 to November 9, 2009. The top subplot graphs the VIX index levels (normalized to percentage by dividing it by 100). There are a total of 5004 daily observations. A noticeable feature is the high levels during the subprime mortgage crisis in years 2007–2009. The bottom subplot graphs the daily VIX changes.

Table 7.3: MLE and PST for ALSEV on the CBOE VIX Index

Panel A: Parametric Specification Test

	κ	θ	δ	L	$\hat{M}$	c_α
Full	96.3726	0.3189	-95.5805	892.45		
Restricted	1.6422	0.2295		1979.40		
					1086.95	730.91

Panel B: Maximum Likelihood Estimation

	κ	θ	δ	$\mathcal{L}$	$2(\mathcal{L}_F - \mathcal{L}_R)$	c_α
Full	3.5332	0.2274	-1.4395	15552.88		
	(0.94)	(0.03)	(1.26)			
Restricted	2.7330	0.2071		15552.13		
	(0.81)	(0.02)				
					1.50	3.84

Table 7.4: MLE and PST for NLSEV on the CBOE VIX Index

Panel A: Parametric Specification Test

	κ	m	δ_1	δ_2	L	$\hat{M}$	c_α
Full	-50.3284	0.2545	0.4584	97.4207	182.94		
Restricted	1.6422	0.2295			1979.40		
						1796.46	877.29

Panel B: Maximum Likelihood Estimation

	κ	m	δ_1	δ_2	$\mathcal{L}$	$2(\mathcal{L}_F - \mathcal{L}_R)$	c_α
Full	-18.2229	0.2438	0.2750	29.5031	15554.99		
	(7.95)	(0.11)	(0.11)	(11.24)			
Restricted	2.7330	0.2071			15552.13		
	(0.81)	(0.02)					
						5.72	5.99

REFERENCES

[1] G. Adimari (1998). An empirical likelihood statistic for quantiles. *J. Statist. Comput. Simul. 60, 85–95.*

[2] Y. Aït-Sahalia (1996a). Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies, 9, 385–426.*

[3] Y. Aït-Sahalia (1996b). Nonparametric Pricing of Interest Rate Derivative Securities. *Econometrica, 64, 527–560.*

[4] Y. Aït-Sahalia, J. Fan and H. Peng (2009). Nonparametric Transition-Based Tests for Jump-Diffusions. *Journal of the American Statistical Association, 104(487), 1102–1116.*

[5] Y. Aït-Sahalia, J. Park (2011). Stationarity-Based Specification Tests for Diffusions When the Process is Nonstationary. Forthcoming in *Journal of Econometrics.*

[6] M. Akritas and I. Van Keilegom (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist. 28, 549-567.*

[7] A. Ang and G. Bekaert (2002). Short Rate Nonlinearities and Regime Switches. *Journal of Economic Dynamics and Control, 26, 1243–1274.*

[8] G. Bakshi, N. Ju and H. Ou-Yang (2005). Estimation of Continuous-Time Models with an Application to Equity Volatility. *Journal of Financial Economics, 82, 227, 49.*

[9] R. Becker, A.E. Clements and A. McClelland (2009). The Jump Component of S&P 500 Volatility and the VIX Index. *Journal of Banking and Finance, 33(6), 1033–1038.*

[10] V. Bentkus and F. Götze (1996). The Berry-Esseen bound for Student's statistic. *Ann. Probab. 24, 491 – 503.*

[11] I. Berkes and L. Horváth (2001). Strong approximation of the empirical process of GARCH sequences. *Ann. Appl. Probab. 11, 789–809.*

[12] I. Berkes and L. Horváth (2003). Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Process. Appl. 105, 271–298.*

[13] I. Berkes, L. Horváth and P. Kokoszka (2003). Asymptotics for GARCH squared residual correlations. *Econometric Theory 19, 515–540.*

[14] R. Bhattacharya and M. Kong (2007). Consistency and asymptotic normality of the estimated effective doses in bioassay. *J. Statist. Plann. and Inference 137, 643–658.*

[15] P. Billingsley (1968). Convergence of Probability Measures. John Wiley & Sons Inc., New York.

[16] N.H. Bingham, C.M. Goldie and J.L. Teugels (1987). Regular Variation. *Cambridge University Press, New York.*

[17] T. Bollerslev (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics 31, 307–327.*

[18] T. Cai, M. Levine and L. Wang (2009). Variance function estimation in multivariate nonparametric regression with fixed design. *J. Multivariate Anal. 100, 126–136.*

[19] T. Cai and L. Wang (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *Ann. Statist. 36, 2025–2054.*

[20] R. Cao and I. Van Keilegom (2006). Empirical likelihood tests for two-sample problems via nonparametric density estimation. *Canadian Journal Statistics, 34, 61–77.*

[21] N.H. Chan, S. Deng, L. Peng and Z. Xia (2007). Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *J. Econometrics 137, 556–576.*

[22] N.H. Chan and S. Ling (2006). Empirical likelihood for GARCH models. *Econometric Theory, 22, 403–428.*

[23] K.C. Chan, G.A. Karolyi, F.A. Longstaff and A.B. Sanders (1992). An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance, 47, 1209–1227.*

[24] N.H. Chan and S. Ling (2006). Empirical likelihood for GARCH models. *Econometric Theory 22, 403–428.*

[25] N.H. Chan, L. Peng and D. Zhang (2010). Empirical likelihood based confidence intervals for conditional variance in heteroscedastic regression models. *Econometric Theory 27, 1–24.*

[26] D.A. Chapman and N.D. Pearson (2000). Is the Short Rate Drift Actually Nonlinear? *Journal of Finance, 55(1), 355–388.*

[27] L.-H. Chen, M.-Y. Cheng and L. Peng (2009). Conditional variance estimation in heteroscedastic regression models. *J. Statist. Plann. and Inference 139, 236-245.*

[28] S. Chen and J. Gao (2007). An adaptive empirical likelihood test for parametric time series regression models. *Journal of Econometrics, 141, 950–972.*

[29] S. Chen and P. Hall (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Annals of Statistics, 21 (3), 1166–1181.*

[30] M. Cheng and L. Peng (2002). Regression modeling for nonparametric estimation of distribution and quantile functions. *Statistica Sinica 12, 1043–1060.*

[31] J. Chen, L. Peng and Y. Zhao (2009). Empirical likelihood based confidence intervals for copulas. *Journal of Multivariate Analysis, 100, 137 – 151.*

[32] S. Chen and Y. Qin (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika 87, 946-953.*

[33] S. Chen and I. Van Keilegom (2009a). A goodness-of-fit test for parametric and semiparametric models in multiresponse regression. *Bernoulli 15, 955–976.*

[34] S. Chen and I. Van Keilegom (2009b). A review on empirical likelihood methods for regression. *TEST 18, 415–447.*

[35] Y.S. Chow and H. Teicher (1988). Probability Theory: Independence, Interchangeability, Martinagles. *Springer-Verlag.*

[36] G. Claeskens, B. Jing, L. Peng and W. Zhou(2003). Empirical likelihood confidence regions for comparison distributions and ROC curves. *The Canadian Journal of Statistics, 31(2), 173 – 190.*

[37] M. Csörgő and L. Horváth (1993). Weighted Approximations in Probability and Statistics. *John Wiley & Sons.*

[38] M. Csörgő and J. Steinebach (1991). On the estimation of the adjustment coefficient in risk theory via intermediate order statistics. *Insurance Math. Econom. 10, 37–50.*

[39] R.B. D'Agostino and M.A. Stephens (1986). Goodness-of-fit-techniques. Dekker.

[40] L. de Haan and T.T. Pereira (1999). Estimating the index of a stable distribution. *Statist. & Probab. Lett. 41, 39 - 55.*

[41] L. de Haan and L. Peng (1997). Slow convergence to normality: an edgeworth expansion without third moment. *Probability and Mathematical Statistics 19 (2), 395 – 406.*

[42] A.L.M. Dekkers and L. de Haan (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist. 17, 1795–1832.*

[43] A.L.M. Dekkers, J.H.J. Einmahl and L. de Haan (1989). A moment estimator for the index of an extreme value distribution. *Ann. Statist. 17, 1833–1855.*

[44] H. Dette, N. Neumeyer and K. Pilz (2005). A note on nonparametric estimation of the effective dose in quantal bioassay. *J. Amer. Statist. Assoc. 100, 503–510.*

[45] H. Dette and S. Volgushev (2008). Non-crossing nonparametric estimates of quantile curves. *J. Roy. Statist. Soc. ser. B 70, 609–627.*

[46] L. de Haan, L. and A. Ferreira (2006). Extreme Value Theory, An Introduction. Springer.

[47] M.D. Donsker (1952). Justification and Extension of Doobs Heuristic Approach to the Komogorov-Smirnov Theorems. *Annals of Mathematical Statistics, 23, 277–281.*

[48] G. Dotsis, D. Psychoyios and G. Skiadopoulos (2007). An Empirical Comparison of Continuous-Time Models of Implied Volatility Indices. *Journal of Banking and Finance, 31(12), 3584–3603.*

[49] G. Durham (2003). Likelihood-Based Specification Analysis of Continuous-Time models of the Short-Term Interest Rate. *Journal of Financial Economics, 70(3), 463–487.*

[50] J. Einmahl and I. Van Keilegom (2008). Specification tests in nonparametric regression. *J. Econometrics, 143, 88-102.*

[51] R.F. Engle (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50, 987–1007.*

[52] J. Fan and I. Gijbels (1996). Local Polynomial Modelling and Its Applications. *Chapman & Hall, London.*

[53] J. Fan and Q. Yao (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika 85, 645-660.*

[54] W. Feller (1971). An Introduction to Probability Theory and Its Applications, Vol. II. *Wiley.*

[55] E. Giné, F. Götze and D. Mason (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab. 25, 1514 – 1531.*

[56] Y. Gong, Z. Li and L. Peng (2009). Empirical likelihood intervals for conditional Value-at-Risk in ARCH/GARCH models. *Journal of Time Series Analysis, 31, 65–75.*

[57] Y. Gong, M. Li and L. Peng (2011). A specification test for nested stochastic models with discrete and dependent observations. Technical Report.

[58] Y. Gong and L. Peng (2010). Coverage accuracy for a mean without third moment. *Journal of Statistical Planning and Inference, 104, 1082–1088.*

[59] Y. Gong, L. Peng and Y. Qi (2010). Jackknife empirical likelihood method for ROC curve. *Journal of Multivariate Analysis, 101, 1520–1531.*

[60] C. Gouriéroux (1997). ARCH models and financial applications. *Springer Series in Statistics.*

[61] P.F. Griner, R.J. Mayewski, A.I. Mushlin and P. Greenland (1981). Selection and interpretation of diagnostic tests and procedures. *Annals of Internal Medicine, 94, 555–600.*

[62] J. Gu, S. Ghosal and A. Roy (2008). Bayesian bootstrap estimation of ROC curve. *Statistics in Medicine, 27, 5407–5420.*

[63] P. Guggenberger and R.J. Smith (2008). Generalized empirical likelihood tests in time series models with potential identification failure. *Journal of Econometrics, 142, 134–161.*

[64] P. Hall (1987). Edgeworth expansion for Student's t statistic under minimal moment conditions. *Ann. Probab. 15, 920 – 931.*

[65] P. Hall (1990). The Bootstrap and Edgeworth Expansion. *Springer.*

[66] P. Hall, R.J. Hyndman and Y. Fan(2004). Nonparametric confidence intervals for receiver operating characteristic curves. *Biometrika, 91, 743–750.*

[67] P.Hall and Q. Wang (2004). Exact convergence rate and leading term in central limit theorem for student's t statistic. *Ann. Probab. 32(2), 1419 – 1437.*

[68] P. Hall and Q. Yao (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica 71, 285–317.*

[69] J.D. Hamilton (1994). Time Series Analysis. Princeton University Press.

[70] L.P. Hansen and J.A. Scheinkman (1995). Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes. *Econometrica, 63(4), 767–804.*

[71] Y. Hong and H. Li (2005). Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates. *Review of Financial Studies, 18 (1), 37–84.*

[72] L. Horváth and P. Kokoszka (2001). Large sample distribution of weighted sums of ARCH(p) squared residual correlations. *Econometric Theory 17, 283–295.*

[73] L. Horváth, P. Kokoszka and G. Teyssière (2001). Empirical process of the squared residuals of an ARCH sequence. *Ann. Statist. 29, 445–469.*

[74] F. Hsieh and B.W. Turnbull (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics, 24, 25-40.*

[75] G.J. Jiang and R.C.A. Oomen (2008). Testing for Jumps when Asset Prices are Observed with Noise – A "Swap Variance" Approach. *Journal of Econometrics, 144(2), 352–370.*

[76] B. Jing, J. Yuan and W. Zhou (2009). Jackknife empirical likelihood. *Journal of American Statistical Association, 104(487), 1224-1232.*

[77] C.S. Jones (2003). Nonlinear Mean Reversion in the Short-Term Interest Rate. *Review of Financial Studies, 16, 793–843.*

[78] P. Jorion (2006). Value at Risk: The New Benchmark for Managing Financial Risk, 3rd ed. *McGraw-Hill.*

[79] A. Keziou and S. Leoni-Aubin (2008). On empirical likelihood for semiparametric two-sample density ratio models. *Journal of Statistical Planning and Inference, 138, 915–928.*

[80] M. Li (2010). A Damped Diffusion Framework for Financial Modeling and Closed-form Maximum Likelihood Estimation. *Journal of Economic Dynamics and Control, 34, 132–157.*

[81] Z. Li, Y. Gong, and L. Peng (2011). Empirical likelihood intervals for conditional Value-at-Risk in heteroscedastic regression models. *Scandinavian Journal of Statistics.* Accepted.

[82] Z. Li, Y. Gong, and L. Peng (2010). Empirical likelihood methods for intermediate Quantiles. *Statistics and Probability Letters 80, 1022–1029.*

[83] M. Li, N.D. Pearson and A.M. Poteshman (2004). Conditional Estimation of Diffusion Processes. *Journal of Financial Economics, 74, 31–66.*

[84] S. Ling (2007). Self-weighted and local quasi-maximum likelihood estimator for ARMA-GARCH/IGARCH models. *Journal of Econometrics 140, 849–873.*

[85] Y. Liu, C. Zou and R. Zhang (2008). Empirical likelihood for the two-sample mean problem. *Statistics & Probability Letters, 78, 548–556.*

[86] C.J. Lloyd (1998). The use of smoothed ROC curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association, 93, 1356–1364.*

[87] C.J. Lloyd and Z. Yong (1999). Kernel estimators of the ROC curve are better than empirical. *Statistics & Probability Letters, 44, 221–228.*

[88] R.M. Kosorok (2008). Introduction to Empirical Processes and Semiparametric Inference. Springer Series in Statistics.

[89] A.J. McNeil, R. Frey and P. Embrechts (2005). Quantitative Risk Management, Concepts, Technique and Tools. *Princeton University Press.*

[90] C.Z. Metz (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8, 283–298.*

[91] E. Molanes-Lopez, I. Van Keilegom and N. Veraverbeke (2009). Empirical likelihood for non-smooth criterion functions. *Scand. J. Statist., 36, 413–432.*

[92] J.P. Morgan (1996). RiskMetrics Technical Document. 3rd ed. *J.P. Morgan, New York.*

[93] H.G. Müller and T. Schmidt (1988). Kernel and probit estimates in quantal bioassay. *J. Amer. Statist. Assoc. 83, 750–759.*

[94] N. Neumeyer and I. Van Keilegom (2010). Estimating the error distribution in nonparametric multiple regression with application to model testing. *J. Multivariate Anal. 101, 1067–1078.*

[95] D.J. Nordman and S.N. Lahiri (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Annals of Statistics, 34, 3019–3050.*

[96] D.J. Nordman, P. Sibbertsen and S.N. Lahiri (2007). Empirical likelihood confidence intervals for the mean of a long-range dependent process. *Journal of Time Series Analysis, 28, 576–599.*

[97] T. Otsu (2006). Generalized empirical likelihood inference for nonlinear and time series models under weak identification. *Econometric Theory, 22, 513–527.*

[98] A. Owen (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75, 237-249.*

[99] A. Owen (1990). Empirical likelihood ratio confidence regions. *Ann. Statist. 18, 90-120.*

[100] A. Owen (2001). Empirical Likelihood. *Chapman & Hall, London.*

[101] D. Park and S. Park (2006). Parametric and nonparametric estimators of the ED100 α. *J. Statist. Comp. Simu. 76, 661–672.*

[102] L. Peng and X. Zhou (2004). Local linear smoothing of receiver operator characteristic (ROC) curves. *Journal of Statistical Planning and Inference, 118, 129 – 143.*

[103] L. Peng and J. Yang (2009). Jackknife method for intermediate quantiles. *Journal of Statistical Planning and Inference 139, 2372–2381.*

[104] J. Pickands III (1975). Statistical inference using extreme order statistics. *Ann. Statist. 15, 1174–1207.*

[105] J. Qin and J. Lawless (1994). Empirical likelihood and general estimating equations. *Ann. Statist. 22, 300-325.*

[106] G.S. Qin and X.H. Zhou (2006). Empirical likelihood inference for the area under the ROC curve. *Biometrics, 62, 613–622.*

[107] J. Ren (2008). Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *Annals of Statistics, 36, 147–166.*

[108] J. Shao and D. Tu (1995). *The Jackknife and Bootstrap. Springer.*

[109] J. Shen and S. He (2007). Empirical likelihood for the difference of quantiles under censorship. *Statistical Papers, 48, 437–457.*

[110] S.E. Shreve (2004). Stochastic Calculus for Finance II: Continuous-Time Models. Springer.

[111] B.W. Silverman (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics Vol. 6, No. 1, 177–184.*

[112] B.W. Silverman (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall/CRC.

[113] H. Taniai and M. Taniguchi (2008). Statistical estimation errors of VaR under ARCH returns. *J. Statist. Plann. and Inference 138, 3568–3577.*

[114] H. Takamizawa (2008). Is Nonlinear Drift Implied by the Short End of the Term Structure? *Review of Financial Studies, 21(1), 311–346.*

[115] T. Tervirta (2009). An Introduction to Univariate GARCH Models. In: Mikosch, T., Kreiss, T., Davis, J.-P., Andersen, R. A., Gustav, T. (Eds.), Handbook of Financial Time Series. Springer, pp. 17–42.

[116] R.S. Tsay (2002). Analysis of Financial Time Series. *Wiley Series in Probability and Mathematical Statistics. Wiley, New York.*

[117] O. Vasicek (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics, 5, 177–188.*

[118] L. Viharos (1997). Tail index estimation based on linear combinations of intermediate order statistics. *Statistica Neerlandica 51, 164–177.*

[119] M.P. Wand and M.C. Jones (1994). Kernel Smoothing. Chapman and Hall/CRC.

[120] Q. Wang and B.-Y. Jing (1999). An exponential nonuniform Berry-Esseen bound for self-normalized sums. *Ann. Probab. 27, 2068 – 2088.*

[121] Q. Wang, B.-Y. Jing and L.C. Zhao (2000). The Berry-Esseen bound for Studentized statistics. *Ann. Probab. 28, 511 – 535.*

[122] L. Wang, L.D. Brown, T. Cai and M. Levine (2008). Effect of mean on variance function estimation in nonparametric regression. *Ann. Statist. 36, 646–664.*

[123] L. Xue and L. Zhu (2006). Empirical likelihood for single-index models. *Journal of Multivariate Analysis, 97, 1295–1312.*

[124] L. Xue and L. Zhu (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of American Statistical Association, 102, 642–654.*

[125] J. You, G. Chen and Y. Zhou (2006). Block empirical likelihood for longitudinal partially linear regression models. *Canadian Journal of Statistics, 34, 79–96.*

[126] K. Yu and M.C. Jones (2004). Likelihood-based local linear estimation of the conditional variance function. *J. Amer. Statist. Assoc. 99, 139–144.*

[127] Y. Zhao and F. Chen (2008). Empirical likelihood inference for censored median regression model via nonparametric kernel estimation. *Journal of Multivariate Analysis, 99, 215–231.*

[128] Y. Zhao and W. Jian (2007). Analysis of longitudinal data in the case-control studies via empirical likelihood. *Communications in Statistics–Simulation and Computation, 36, 565–578.*

[129] Y. Zhao and S. Yang (2008). Empirical likelihood inference for censored median regression with weighted empirical hazard functions. *Annals of the Institute of Statistical Mathematics, 60, 441–457.*

[130] Y. Zhou (1996). The Rates of Convergence of the Oscillation Modulus for Empirical Process under Dependence. *Chinese Annals of Mathematics, vol. 17, ser. A No.1, 1–8.*

[131] K.H. Zhou, W.J. Hall and D.E. Shapiro (1997). Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine, 16, 2143–2156.*

[132] Y. Zhou and H. Liang (2005). Empirical-likelihood-based semiparametric inference for the treatment effect in the two-sample problem with censoring. *Biometrika, 92, 271–282.*

[133] W. Zhou (2008). Statistical inference for $P(X < Y)$. *Statistics in Medicine, 27, 257–279.*

[134] M.H. Zweig and G. Campbell (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39, 561–577.*